

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

<http://go.warwick.ac.uk/wrap/76664>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

**The Authority of Us:**  
**On the Concept of Legitimacy and the Social Ontology of Authority**

by

Adam Robert Arnold

A thesis submitted in partial fulfilment of the requirements of the degree  
of  
Doctor of Philosophy in Philosophy

University of Warwick, Department of Philosophy  
October, 2015

## Contents

Acknowledgements.....	iii
Declaration.....	iv
Abstract.....	v
Introduction: Social Institutions, Social Standpoints and Social Reasons.....	1
Part I: On the Concept of Legitimacy: Two Questions for Authority.....	11
One: The Concept of Authority.....	12
1. Reformulating the Question of Authority.....	15
2. The Question of Standing and the Question of Content.....	21
3. Having Authority: Epistemic vs. Practical Authority.....	26
4. Normative Magic and the Mystical Foundations of Authority.....	32
Two: Two Individualist Conceptions of Authority.....	36
1. The Service Conception of Authority: An Answer to the Question of Content.....	36
2. Does the Normal Justification Thesis Deliver Command Based Pre-Emptive Reasons?.....	41
3. The Second-Person Standpoint in General.....	46
4. Second-Person Standpoint and Practical Authority: An Answer to the Question of Standing.....	50
5. Coercion, Accountability and the Right to Rule.....	55
6. Radicalising the Question of Standing: A Critique of Darwall's Second Person Standpoint.....	64
Three: Authority as Social Practice.....	72
1. Modest Sociality and Social Practices: The Inadequacy of Sociality Based on Content.....	73
2. The persistence Interdependence of Shared Agency and the Problem of Instability	77
3. Normative Construction of Agency, Practical Identity and the Bindingness of Commitments.....	89
4. Own-Action Condition and the Mineness of Intentions.....	100
5. Robust Sociality: Grounding Authority in Social Practice.....	104
Part II: Governing Together: On the Social Ontology of Authority.....	106
Four: On the Bindingness of Social Ontology.....	107
1. The Social Ontology of Institutions.....	108
2. The Problem of Collectivity: The Necessity for a Social Ontology of Institutions	111
3. The Bindingness of Institutions.....	117

4. The Constitution of Social Practices: We-Mode Accounts and Subject Accounts..	123
Five: Social Ontology and the Question of Standing: The Owing Account.....	126
1. Standing and Authority as Owing Relation.....	129
2. The Social Ontology of Authority: Joint Commitment as Foundation.....	131
3. The Authority of Us.....	145
Six: Social Ontology and the Question of Content: Unity through Social Reason.....	148
1. Constitutive Commitments.....	151
2. What are Social Reasons?.....	160
3. Proceduralism and Content.....	165
4. Directed Obligations: Social Standpoints and the Tragedy of the Social.....	175
Bibliography.....	193

## Acknowledgements

First and foremost I need to express my deepest gratitude to Fabienne Peter. Without her guidance and support throughout the writing of this thesis it would have never been completed. I would also like to express my gratitude to my examiners: Margaret Gilbert and Peter Poellner. Their thoughtful, challenging and deep questioning of my work was a deeply edifying experience. Kimberly Brownlee is owed a lot for arranging and chairing my viva. I owe a large part of my intellectual and philosophical development to the academic staff of the Warwick Philosophy Department both past and present. Particularly I would like to thank: Professor Keith Ansell-Pearson, Professor Stephen Butterfill, Professor Christoph Hoerl, Professor Stephen Houlgate, Professor Massimo Renzo, Dr. Robert Cowan. The graduate community of the philosophy department was a hugely stimulating environment and I must thank the following members for many challenging and thoughtful conversations about my work in particular and philosophical topics in general: Jamie Abernethy, Alfonso Anaya Ruiz Esparza, Benjamin Berger, Wesley Chai Sam Clarke, Joesph Cunningham, Matthew Dennis, Juan Camilo Espejo-Serna, Peter Fossey, Tania Ganitsky, Ivan Ivanov, Dino Jakušić, Richard Lambert, Justin Neville-Kaushall, Bethany Parsons, David Rowthorn, Roberta Locatelli, Irina Schumski, William Stafford, Florence Sunnen, Nicola Spinelli, Daniel Vanello, Barnaby Walker, Graham Wetherall. A thank you is also deserved by all members of Warwick's Centre for Ethics, Law and Public Affairs. This work has also been greatly improved by the audiences at the following conferences: Collective Intentionality IX, Collectivity Conference, and Thinking (about) Groups. A special thank you is owed to Professor Amos Nascimento whose support and encouragement throughout the years has been deeply important to me. Finally, I would like to thank my friends without whose support I could have never made it this far. This work is dedicated to my mother, Niki Olson.

## **Declaration**

This thesis is my own work and contains no material submitted before for any other degree.

## **Abstract**

Authority figures permeate our daily lives, particularly, our political lives. What makes authority legitimate? The current debates about the legitimacy of authority are characterised by two opposing strategies. The first establish the legitimacy of authority on the basis of the content of the authority's command. That is, if the content of the commands meet some independent normative standard then they are legitimate. However, there have been many recent criticisms of this strategy which focus on a particular shortcoming – namely, its seeming inability to account for who can legitimately command whom. This is the basis of the second strategy, which attempts to characterize the normative relationship that underlies and makes possible authoritative commands. The central point of Part I is that these two strategies are, in fact, not opposed and both raise questions which a theory of legitimacy must answer. If this is the case, then we need to ask: how ought we to determine the legitimacy both of the content of commands as well as who can command whom? Part II will answer this question. Starting with the question of standing, I argue that we ought not to look for normative principles outside of the institutions in which authority is embedded. Rather, one ought to start by elaborating the ontology of institutions in which a *sui generis* form of normativity arises. A joint commitment account of social ontology provides the tools necessary to see how the direction obligations emerge concurrently with the formation of institutions. Similarly, the question of content can be answered by paying close attention to the social ontology of institutions. We need not look beyond the internal constitutive standards of the institution itself. The constitutive standards provide an internal criterion by which the legitimacy of commands can be established.

## **Introduction: Social Institutions, Social Standpoints and Social Reasons**

Authority figures permeate our daily lives and, in particular, our political lives. Political offices are paradigm cases of practical authority. These authorities claim to have the ability to command us to act in a certain way. They are practical authorities because they are concerned with action, as opposed to epistemic authorities who are concerned with what we ought to believe. When practical authorities issue commands, it is normally presumed that they have the right to issue these commands; that is, authorities have a right to rule even if it is in a limited domain. Concurrently, it is ordinarily presumed that the addressees ought to defer to this authority and that there is, internal to the command itself, a demand on the addressees to conform.

These types of political office are both prominent and troubling for two particular reasons. First, we seem not to have much of a choice in being part of a political society, or a state (Hume 1985, pp. 475-476 and Dunn 1996, p. 66). Even if we are fortunate enough to have the means to leave the particular state into which we are thrown at birth, we would most likely leave it only for another state. Second, the authoritative commands of states are backed by coercive power – they claim a “monopoly of the legitimate use of physical force in the enforcement of [their] order[s]” (Weber 1968, p. 54; cf. Weber 1946, p. 78).

What makes these authorities legitimate, and when are authoritative commands justified? The term ‘authoritative commands’ may seem pleonastic, and in a sense it is. However, I will use the term ‘authoritative commands’ to denote ‘commands which are given by a legitimate authority’, as opposed to ‘commands given by a purported authority’. The term ‘authoritative commands’ is then to be taken as a success term.



In the following work, I will argue that the best way to understand the legitimacy of authority is to understand how authorities are embedded in institutions. The idea to be explored is that when authorities command their addressees, they give the addressees new reasons for action (this will be further discussed in Chapter One). However, these reasons for action are not moral reasons for actions but social reasons for action. They are social reasons because, unlike moral reasons, these reasons are limited only to those who belong to the institution in which the authority is embedded.

How will an institutional model of authority help with the question of legitimacy? The validity of the social reasons given by the commands is determined by the role that the authority plays in the institution and what role the authority is empowered by the institution. In order to see how this works, we need to understand better what an institution is. In particular, we need to understand how the structure of an institution contains its own form of normativity internal to the institution.

“Two Concepts of Rules”, an early essay by John Rawls, points clearly to this form of normativity insofar as he defines ‘practice’ as:

a sort of technical term meaning any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and which gives the activity its structure (Rawls 1999b, p. 20n1).

In later works, Rawls favours the word ‘institution’ over that of ‘practice’:

Now by an institution I shall understand a public system of rules which defines offices and positions with their rights and duties, powers and immunities, and the like. These rules specify certain forms of action as permissible, others as forbidden; and they provide for certain penalties and defenses, and so on, when violations occur (Rawls 1999a, pp. 47-78; cf. Thompson 2012, p. 193)

In turn, I too will favour the term ‘institution’. Rawls deploys this idea of institution in order to show the importance of the distinction between justifying an institution and justifying a particular act falling under the institution. On the one hand, justifying an

institution involves answering questions such as: Why this is institution important? Why is it structured the way it is? Why these rules and not some other rules? On the other hand, to justifying a particular act falling under an institution, we would ask: Does this act count as an appropriate act or move? The latter question clearly points to a form of normativity. Hence, an inappropriate move is a move that is not allowed within a particular institution.

One of the examples that Rawls uses to highlight this distinction is baseball (Rawls 1999b, pp. 37f). As he describes the game, there are several actions one could perform alone or with others, such as hitting, running and throwing. However, these actions do not count as playing baseball unless they occur during or inside of a game. There really is no sense to 'striking out' or 'stealing a base' outside of or before one is involved in a game; that is, in the institution of baseball.

We can call this Rawls's priority claim: the institution is logically prior to a particular move. The reason we refer to priority here is that what counts as an appropriate or justified move in the game of baseball necessitates the rules of the game being established prior to the acts involved in the game. For example, in order for stealing a base to be a legitimate move in the game of baseball, there must be rules which establish such a move as a legal possibility. Understanding the logical priority of rules to acts is what grounds Rawls's distinction.

In justifying the institution of baseball we could ask questions such as, 'Why should a player only be allowed three strikes?' These types of questions need to be asked outside of or external to the institution; the reason for this should be easy to understand. If we were trying to play a game of baseball and one raised such a question, then the game would have to be stopped until the question was settled. We would not know how to continue the game without such questions being settled first. Hence, questions about the justification of

institution are questions about the constitution and structure of the institutions themselves.

The logically distinct questions of the appropriateness or inappropriateness of particular moves are then to be asked from inside the institution; for example, the question of whether it is appropriate for the umpire to call a 'strike' or a 'ball'. The meaning of these terms is constituted by the rules of the game and determines what call the umpire should make, not excluding difficulties of vagueness, indeterminacy or what H.L.A. Hart called 'the open texture' of laws/rules for the discretion and judgement of the umpire (Hart 2012, Ch. VII.1).

The role of umpires as a part of the institution of baseball is analogous to the role of many institutions permeating our everyday life. My reference here is to the offices within institutions empowered with the ability to issue commands: Presidents, Prime Ministers, Congressional Representatives, Members of Parliament, employers, supervisors and police officers, referees etc. Notice that these offices only make sense when they are considered to be embedded within broader institutions. What defines and constitutes these particular offices and their role is the structure of the institutions to which they belong. An umpire only makes sense within the institution of a certain game, like baseball. Political offices only make sense embedded within the institution of governance.

Being that practical authorities are roles embedded in institutions, we can follow Rawls by asking two questions of these authorities. First, is this institution and its roles justified? Second, we can also ask whether or not a particular command by someone occupying the role of authority is appropriate; that is, does this particular act (a command) fall under what is allowed within the framework of the institution?

Which of these questions should we be asking when we consider the legitimacy of an authority? On the one hand, should we think of legitimacy in terms of justifying the

institution or, alternatively, should we think of it in terms of justifying a particular act falling within the rules of the institution? In the work ahead, I will argue that we should take the latter question as the question of legitimacy. In contrast, the former question should be thought of in terms of justice.

Here I am making a highly controversial claim, one that will be seen as orthogonal to most contemporary theories of legitimacy. For this reason, I will begin by closely examining the current debate and showing its shortcomings. This move will open up the space for an alternative approach, one which is truer to the phenomenon at hand, as well as containing more than enough critical resources to avoid the pitfalls of ethnocentric norm-based reasoning (O'Neill 2000, p. 22).

The current debate about the legitimacy of practical authority is characterised by two seemingly opposed strategies. The first strategy to establish the legitimacy of commands is based on the content of the command itself; commands are legitimate if the content of the commands meets some independent normative standard. We may call this 'the instrumental tradition of practical authority' expressing, for example, the strategy of Joseph Raz's highly influential 'service conception of authority' which we will examine in more depth in Chapter Two (Raz 1986, 2009). The instrumental tradition asks: does the content of this political authority's commands advance the interest of the individual more than his being in a hypothetical state of nature, i.e., in a state without the authority? What the authors in this tradition hold in common is their belief that authority is justified if they make individuals better off in comparison to the situation that would result from the absence of the authority. We can therefore understand the meaning of 'better off' as a part of the substantive issue which provides a distinction between particular accounts.

However, there have been many recent criticisms of Raz's account which focus on a

particular shortcoming; namely, Raz's seeming inability to account for who can legitimately command whom. This is the basis of the second strategy, an alternative strategy dominated by the consent tradition, which attempts to characterise the normative relationship that underlies and makes possible authoritative commands.

Those of the consent tradition, such as A. John Simmons, draw inspiration from paragraph 222 of John Locke's *Second Treatise of Government* which also compares the consequences of being subject to a political authority with a situation in which there is no political authority. However, what the consent tradition takes the instrumental tradition to answer is the justification of the state's existence, not the legitimacy of a state's commands (Simmons 2001, Ch. 7). The consent tradition would thus want to ask a further question, looking beyond whether the state, or authorities more generally, are morally defensible.\*

This question that the consent tradition wishes to pose is: What gives this particular authority a right to rule over someone in particular? It is a question that is to be answered by the consent tradition, with a story about an individual's 'positive engagement' with the authority establishing a normative relationship (Locke 2003, ¶122). The thought is that, by properly characterising the normative relationship, we can explain how authoritative commands are possible. This normative relationship is called 'standing' and consists of answering the question of who has the standing to command whom? Stephen Darwall has adopted this strategy in his second-person account of practical authority which will be further discussed in Chapter Two (Darwall, 2013a).

The central point of Part I is that these two strategies are, in fact, not necessarily opposed, with both raising questions that a theory of legitimacy must answer. If this is the

\* Notice that we could make the Hobbesian minimum condition and the consent tradition condition both necessary conditions for legitimacy in a pluralistic account. What is important now is seeing the two major accounts in a pure ideal type account.

case, then we should ask: How ought we to determine the justified content of commands as well as who has the standing to command whom?

Both conceptions of the instrumental tradition and the consent tradition clearly wish to look outside of the institution in order to answer the question of legitimacy. They both want to justify the institution of authority as such; one looking towards the content of the commands, the other looking towards the standing of who can command whom. Furthermore, in both cases we encounter individualistic answers where the evaluation of authority is always from the individual standpoint; that is, from a point of view which is, by necessity, outside of the institution. The maintaining of an individualistic standpoint explains their inability to sustain the separation between questions of legitimacy and questions of justice.

Part II of this thesis will develop a radically different strategy. Starting with the question of standing, I argue that we ought not to start by looking for normative principles independent of the institutions in which authority is embedded. We ought rather to start by elaborating the ontology of institutions in which, internal to the social ontology of such institutions, a *sui generis* form of normativity arises. This *sui generis* form of normativity consists of a new type of reasons: social reasons, referred to by others as 'public reasons' (Freeman 1990b, Korsgaard 1996b, O'Neill 2000). They also have much in common with what Charles Taylor has termed 'common understanding' (Taylor 1995). Following Anthony Simon Laden (Laden 2005), I have chosen to opt for the term 'social reason' to emphasise both its connection to and emergence from the social ontology of institutions.\* As Hegel would say, these reasons apply to individuals *as members* of an institution (Hegel 2008, §158).

\* In opposition to Raimo Tuomela's account, the important type of reasons here are justificatory reasons, not explanatory reasons (Tuomela 2013, p. 99).

When one becomes a member of an institution, one also gains a new practical identity (Korsgaard 1996a, p. 101). This is precisely what it means to be a member of an institution. For example, as a member of a state one gains the practical identity of being a citizen. This practical identity of being a citizen brings with it social reasons to act in certain ways, e.g., to obey the law (Korsgaard 1996a, pp. 102-107, 120, 137f).

The account of social ontology elaborated here is inspired by the work of Margaret Gilbert. Her plural subject view provides the necessary tools to see how the obligations of an owing relationship is constitutive of institutions. It is also this owing relationship that establishes who can legitimately command whom. In a seeming paradox, it is the authority of a 'we' who can command an 'us'. The second chapter of Part II can be seen as an elaboration of the following comment on the part of Gilbert:

one interesting aspect of this conception is that it helps to explain the sense one might have that the source of political authority in every case lies with 'the people'. Here, a joint commitment of the whole population in question – the people – is taken to underlie whichever kind of rule is in place (Gilbert 2006b, p.213).

The resulting conception of political authority, and practical authority more generally, can be traced back through the work of some of the most distinguished figures in the Western philosophical tradition: from Rousseau, Kant and Hegel to the later work of Rawls, Habermas and Korsgaard.

What is at stake is precisely the constitution of social reasons. It is through this idea that the second question that I posed can then be answered. This is the subject of the third chapter of Part II. My main claim in this chapter is that the question of content can be answered through paying close attention to the ontology of institutions. As opposed to seeking normative principles outside of the institution, we have no need, initially to look further than the internal constitutive standards of the institution itself. The constitutive

standards provide an internal criterion by which the legitimacy of commands can be established.

Remaining within institutions in order to understand the legitimacy of authority involves, within the debate about legitimacy, a shift from the standard question. However, it does not follow from this that I am rejecting external points of view and the question of justifying an institution as such. Rather, my point is that the internal, social standpoint of an institution has its own normative force that can be exploited in order to determine the legitimacy of a command.

My point can be illustrated with the following analogous case – the social practice of language.\* Before even asking if someone's claim that  $P$  is true or false (which I will take as structurally similar to the claim that  $\phi$ -ing is morally right or wrong), we have to understand the practice of making this sort of claim; the appropriateness/inappropriateness of claiming  $P$ . For example, one cannot appropriately claim that both  $P$  and  $\neg P$  are true. To claim that  $P \wedge \neg P$  is flatly contradictory (leaving aside controversial claims about the possibility of dialetheic sentences and self-referentiality). Violating the principle of non-contradiction, the “most certain of all principles” (Aristotle 1984a, 1005b24), would make one “no better than a mere plant” according to Aristotle (ibid, 1006a15). The point is that there are some claims which are not raised to the level of appropriate claims. We are therefore not required to ask the further question as to the truth of said claim. This is all internal to the social practice of language.

More in line with the issues at stake here, we can look at the case of the appropriateness or inappropriateness of a proposed law. Before we even need to move to

\* This approach to the philosophy of language has been developed by Robert Brandom starting with his 1979 paper “Freedom and Constraint by Norms” and received its fullest treatment in 1998 with his *Making it Explicit*.



the external point of view and ask if this law is moral or if the law is good for me, we have an internal question to answer: Is this law constitutional (Hart 2012, p. 120)? If the law does not meet this internal standard, then we have no need to ask the further question. Therefore, the question to follow will be: When is it internally appropriate for  $A$  to command  $B$  to  $\phi$ ? I will argue that this is the best way to understand the legitimacy of authority.

Of course, it is reasonable to think that an institution must be just, at least minimally, in order for it to be able to give authoritative commands. If not, the social reasons might be so easy to override that it will not make the least bit of practical difference. For example, a constitution which allows for slavery could not possibly be authorised to enforce slave-holding because the social reasons to remain a slave within that society would be completely over-ridden by moral considerations. This assertion is reasonable and more will be said about this minimally just requirement for practical authorities. However, it does not follow from this minimal requirement that social reasons never make a difference. There will still be innumerable cases in which social institutions meet this minimal standard but fail to be completely just. It is these cases that are of paramount importance to our understanding of the issues at stake. In other words, how do we make sense of a minimally just institution's ability to give reasons to their members?

The fundamental claim, then, is that the legitimacy of practical authority turns on understanding the social ontology that both underlies and makes possible commands in the first place. In other words, the best way to make sense of practical authority is by understanding it to be embedded in an institution, so understating how a particular command is enabled by the rules of the institution. A proper understanding of institution involves giving an account of the social ontology of an institution.

## **Part I: On the Concept of Legitimacy: Two Questions for Authority**

## **One: *The Concept of Authority***

A question which has interested philosophers since at least Plato's *Crito* is: When, if ever, does an alleged authority actually have the right to rule? That is, when do commands legitimately demand conformity from their addressees?

However, this claim is in some ways anachronistic. The concept of authority has its origins in Rome, not Athens, and from the term *auctor* (master, leader, author). In contrast, the Ancient Greeks did not have a single word to capture the complexity of authority (Arendt 2000, p. 473; Geuss 2014, Ch. 6; Agamben 2005, Ch. 6). However, as Hannah Arendt pointed out:

the philosophy of Plato and Aristotle, [...] in quite different ways but from the same political experiences, tried to introduce something akin to authority into the public life of the Greek polis. (Arendt 2000, p. 473)

We could align the question more with what Plato actually asked: Will Socrates act justly if he disobeys the will of the Athenians? (Plato 1997a, 48c). Plato's arguments are usually taken to focus on Socrates's obligation to defer to the commands of the city. However, it is evident that each argument also relies on an implicit argument concerning that which makes authority legitimate. That is, the structure of Plato's argument is that if there is a legitimate authority, then it follows that the addressees of that authority have an obligation to defer by suspending their judgement and conforming their will. Conversely, the addressees will only have this obligation to defer if the authority is legitimate; a fact that becomes evident when we think about how Plato poses the question over and over in the dialogue. Socrates continually takes the point of view of the Athenians and asks what the Athenians would say to him if he had disobeyed.

This argument comes out in the opening exchange in the *Crito*. Crito comes to save

Socrates's life by taking him away from Athens. He thinks that the Athenians are being unjust in commanding Socrates to drink the hemlock. As Crito makes clear, if Socrates follows the commands he will give “up [his] life when [he] can save it” (ibid., 45c). Yet this, for Crito, is an injustice in itself. Furthermore, Socrates would be “betraying [his] sons by going away and leaving them, when [he] could bring them up and educate them” (ibid., 45d; cf. 50c). What Crito is questioning is the legitimacy of the Athenian authorities to demand such things of Socrates.

It would then seem that authoritative commands have a dual nature. When an authoritative command is given, it is ordinarily presumed that the authority has a right to give such a command and the addressee of the command ought to defer to the authority. It is difficult to maintain a separation between these two aspects of commands: the legitimacy of authorities and the obligation of addressees to defer to them. Firstly, this is because authoritative commands are always given by someone in authority to an addressee. Secondly, part of what authorities are claiming for themselves is that their addressees ought to suspend their judgement about what to do and conform their will to the authority's commands.

Further, it is difficult to make sense of the idea that authorities can legitimately make demands on their addressees without the addressees having an obligation. One way to bring this out is to emphasise another ability that authorities claim for themselves, viz., the ability to enforce their commands with coercive power. It is at least, *prima facie*, reasonable to think that at least one necessary condition for legitimising this ability of authority is to show that addressees have, in fact, an obligation to defer. We ought to keep this whole complex in view when thinking about practical authorities. However, different conceptions of authority will surely come down on different sides of this issue. It will not be until we

take a closer look at two major conceptions of authority in Chapter Two that we will be able to resolve these issues.

The arguments that Socrates proceeds to give on behalf of the Athenian authorities are meant to show that they are indeed justified in demanding his compliance. This points to a particular aspect of our question that we should keep in mind, for the criterion we are seeking in answering Plato's question is something less than justice (cf. Rawls 2005, p. 428). He is not asking if the commands are just commands in themselves but rather whether the commands bring with them some obligation to obey. He might even agree with Crito that they are not just. Yet they may still be justified in the sense that he ought to defer to them.

Part of the reason that Socrates gives for thinking that legitimacy is something less than justice can be expressed the simple proverb: two wrongs do not make a right (Plato 1997a, 49a-d). In other words, if it is wrong to disobey the commands of a legitimate authority then it is not enough to show that a command is unjust. Here again, we see the importance of the connection between legitimacy and obligation. When authorities are not legitimate, there are no command-based reasons to defer. There may still be other important reasons to act in conformity to what authorities demand, it is just that the reason does not come from the command itself.

Socrates seems to be making the point that we should not ground our answer to the question of legitimacy in a comprehensive theory of justice. However, this does not rule out there being some overlap between the two. It seems reasonable to think that a completely unjust authority will also fail to be legitimate. This overlap may have to do with the idea that a completely unjust authority will tend to be unpredictable in terms of what they will command and not “let scruples concerning the rule of law interfere with their interests in

particular cases” (Rawls 1999a, p. 52). In other words, the unjust authority will tend only to give commands that further the authorities’ own self-interest.

This appears to be the case with slave-holders and a constitution which allows for slave-holding. The interest of those enslaved are completely subordinated to the arbitrary will of the slave-holder's interests. This is what Kant called barbarism: “force without freedom and law” (Kant 2007, 7:331; Ripstein 2010, p. 336f). As we will see, authorities who in no way take account of the interest of their addressees are unable to give authoritative commands.

Remarkably, Plato was able to outline several of the answers to his question, as he did with so many other philosophical questions that remain prominent today. First, I want briefly to canvas the contemporary arguments for two reasons. First, as with most Platonic dialogues, we are left with unsatisfying answers. One gets the same sense when reading the contemporary literature on legitimacy and political obligation; with all the ingenuity of the different views defended, these answers often end up being one-sided. Second, the one-sidedness of the contemporary arguments points to posing the question of authority in a more nuanced, fuller manner. It is of primary importance to clarify what it is we are asking when we approach a philosophical problem. This is the primary aim of this opening chapter.

## 1. Reformulating the Question of Authority

Let us begin with the argument from fair play (Hart 1955, Rawls, 1999b Ch. 6 and Klosko 1987). The main thought behind this argument is that when one participates and benefits in a cooperative joint enterprise, one also incurs an obligation to do one's fair share to support the cooperative joint enterprise. What is implicit in this argument is the idea that

the authority is legitimate because of the beneficial consequences it provides to its subjects through supporting or making possible the cooperative enterprise. This fundamental line of thought - that authorities are justified through the benefits they are able to provide - carries through to the next arguments to be considered. These arguments then fit broadly in the instrumental tradition outlined in the introduction. There is an implicit comparison here between the life in the cooperative joint enterprise and how life would be in a hypothetical state of nature.

Another similar type of argument comes from utilitarianism. The argument here is that, if deferring to the commands of authorities is for the discretion of the individual, then commands will have no force whatsoever and the unifying function of the authority will no longer be effective. Similar to the last argument, the idea seems to be that people are better off with authorities than existing in the state of nature. However, John Horton notes that utilitarian arguments tend not to be incorporated into contemporary arguments about political obligation and legitimacy:

Those few theorists who have sought to incorporate a substantial utilitarian component within their justification of political obligation have invariably done so in a highly qualified manner; and there have been few if any attempts to articulate a fully elaborated theory of political obligation in uncompromisingly utilitarian terms. As utilitarianism is a far from new or underdeveloped theory, this is of itself a significant indication of its limitations in this area. (Horton 2010, p. 69 citation removed).

However, we can still see the intuitive force behind this idea. The justification of authority is parasitic on the overall benefit that the addressees as a whole receive from the existence of the authority. If the authorities are of benefit overall to their addressees, then the authority has the right to rule.

Next, following a similar move, we might present the argument from theory of gratitude (Walker 1988). The fundamental premise in this argument is that when one

received benefits from someone one has an obligation not to do anything against the interests of whoever provided the benefits. Moving from this premise, the argument runs in the following way. Since citizens have benefited from their state then they have an obligation not to do anything against the interest of the state. Disobeying the commands of state authorities goes against the interest of the state, for similar reasons as those presented in the utilitarian argument. Therefore, citizens have an obligation to follow that commands of the state.

Finally, there is the natural duty account (Waldron 1993 and Wellman 2005). The basic idea behind this account is that in order for individuals to discharge some duty (e.g., justice, the duty to rescue), the individual must follow the commands of a state. Several distinct reasons are provided here, e.g. under-determination of what is owed or issues of disagreement. But again, this account turns on the authority providing some form of benefit - i.e., allowing their citizens to discharge some duty they already have - as that which grounds its justification.

These first four arguments can be generally seen as arguments following what I have termed 'the instrumental tradition' insofar as they gain their justificatory power from the benefits the individual gains from the authority being in place. In other words, the institution of authority is justified through reference to the individual's interest. They are all meant to show why the content of authoritative commands make the individual better off than in the state of nature. However, what these arguments fail to establish is any sense as to why this applies to some authorities and not others; that is, why are the authorities entitled to make demands on any particular individual in the first place. Or, in the same vein, why does a particular individual not owe their obedience to all the authorities providing these benefits? (Simmons 1979, Ch. II.i)



In other words, we might want to respond to the above arguments by asking why these particular authorities - particularly if other authorities could provide the benefits more effectively - bind us to the particular states to which we belong. Or, why these authorities and not some others?

The obvious response to such a question would be to turn to consent theories (Locke, 1993; cf. Simmons 1979). By agreeing to be part of the state, we become obligated to defer to the commands of the state. This is one of the most intuitive and compelling argument for why one is obligated to defer to authorities. How does this account for legitimacy? The idea is that it is a necessary condition for the legitimacy of authorities that the addressees of that authority have consented. This clearly provides a tie between particular individuals and a particular authority. However, one might suspect that, in reality, few people have ever consented to their state. Further, it is unclear why, if one consents to a particular authority, that one could not withdraw one's consent at any time.

Unlike the instrumental tradition, these arguments do seem to go a long way towards showing why these particular authorities are owed their subjects deferral. Yet they say nothing about the content of the commands that authorities are justified in giving. From the perspective of the instrumental tradition, we may wish to ask 'What kind of normative force do these relationships have if we do not pay attention to their content?' Up to now, many have believed that we cannot agree to be enslaved, so demonstrating that consent, for example, must have some content-ful limits which these relational accounts do not, on their own, provide.

These types of worries have led to the recent development of pluralistic accounts. Two particularly important pluralistic account are the associative account of obligation developed by Ronald Dworkin (1986, Ch. 6) and John Horton (2006, 2007 and 2010). The

idea behind this type of account is to establish the similarity between the special relationship between parents and children; that which brings with it special obligations and duties, just as the special relationship between citizens and their state gives rise to special obligations and duties. The idea is that membership in and of itself entails obligations.

Furthermore, in different ways, both Dworkin's and Horton's accounts attempt consistently to provide conditions to meet the worries of those belonging to the instrumental tradition. For example, one condition Dworkin places on associative obligation is that each member must be concerned with the well-being of others in the group (Dworkin 1986, p. 200). Horton makes a similar move when he claims that associative obligations can only arise from “associations that can have value for its members” (Horton 2010, p. 176). These two accounts attempt to account not only for the beneficial consequences of authority, but also the relationship that ties individuals to their particular authority. The account that will be developed in Part II will be in this pluralistic tradition. Therefore, we can hold off on a further discussion of Dworkin and Horton until then.

However, from this all too brief discussion, I want to suggest that it is helpful to break Plato's fundamental question down into two further questions. First, we should ask: when, if ever, is the content of a command justified? This has been a dominant topic in contemporary political philosophy, largely due to the influence of philosophical anarchism (Raz 1979, ch. 1). The idea is that authorities are legitimate only if the content is justified. The other important question, which has recently been overshadowed, viz., when is one justified in being in a position of authority over another? When is *A* justified in having authority over *B*? The first question I will term 'the question of content' and the second, 'the question of standing'.

The question of standing has gone under different names; for example, the 'mark', 'sign', 'symbol', 'certificate' or 'credentials' of authority (Friedman 1990, p. 70). This latter question, it will now be argued, is not a purely moral question. Rather, it is fundamentally a question of social ontology and the social constitution of authority in which a *sui generis* form of normativity arises. That is, by taking for granted social ontology, theories about authority miss a fundamental aspect about authority; or, in Mark Turner's words, the task of slaying a 'dragon of obliviousness' (Turner 2009), being that that most conceptions of authority pay little heed to the social ontology which binds individuals to their institutions. It is in this way that the answer to this latter question will have fundamental implications for our answering the former question. The implication being that the constitution of the institutions empowers authorities to create social reasons as well as limits these same reasons; this will be the topic of Chapter Six.

To put it differently, we can ask the question: Who has authority over whom and for whose benefit? Here, both questions are brought together in a manner maintaining the importance of both. The formulation is adopted from Raymond Geuss's discussion of Lenin's question of 'Who whom?', although the question of 'Who whom?' can be found as far back as at least Hobbes (Geuss 2008, pp. 23-30; cf. Raz 1979, p. 10). The second part of the question, 'for whose benefit?', can be seen as far back as Cicero's "Pro Roscio Amerino" where he attributes the phrase, *cui bono?*, to Lucius Cassius; here, the context of this question is a trial seeking to discover hidden motives (Cicero 2000, §84 and p.228n84).

The first half, the 'who/whom', refers to 'the question of standing', while the second half, 'for whose benefit', refers to 'the question of content'. It is important to be clear that 'benefit' in this question does not imply a theory tying the question of content to good consequences for the addressee of the command. A deontologist could reasonably claim

that the necessary benefit which authorities must provide is the assurance of freedom. For a deontologist, authorities would only be justified if they provide the benefit of protecting moral borders around the person (Schmidtz 1996, p. 84; Kant 1996c and Ripstein 2009).

I will argue, following Dworkin and Horton, that only if we answer both questions – 'the question of content' and 'the question of standing' – can we determine when authority is legitimate; that is, whether or not an alleged authority in fact has the right to rule. The right to rule should be thought of in terms of the authority being entitled to the addressee's deferral to the authority's commands. The reason for framing the issue in this way is that it highlights a significant point about authoritative relationships. As we shall see, authorities, in the sense at issue here, do not merely hope that their commands are followed but are also entitled to the compliance of their addressees. If the subjects do not comply, then the authorities are also entitled to use force to enforce their commands and to ensure compliance (more on this in Chapter Two). I therefore take it that having the right to rule is what it means to be a legitimate practical authority.

## 2. The Question of Standing and the Question of Content

These two questions, the question of content and the question of standing, can easily be conflated, which can partly be explained by how tightly related the questions are (Friedman 1990). To begin, let us examine the different implications of each question for the legitimacy of authority.

It is at least conceptually possible that someone may have the standing to give commands without giving authoritative commands. Think of a corrupt police officer who terrorises everyone on his beat. Detective Alonzo Harris, Denzel Washington's character from the Antoine Fuqua film *Training Day*, exemplifies such a police officer. He is in debt

to the Russian Mafia and in order to repay his debt, he uses his status and his force to steal drugs and money from those he is meant to protect. Let us assume that Harris has the standing to give authoritative commands, at least those of a limited nature, being he is an active duty police officer. However, it so happens that the commands he does give (for example, using fake search warrants to steal drug money) are illegitimate. Nonetheless, they have an aura of legitimacy which is parasitic on his standing as a police officer.

That the aura of legitimacy may be highly influential on people's behaviour is evident not only throughout history, e.g. the Mỹ Lai Massacre ordered by U.S. Lieutenant William L. Calley on 16 March, 1968 (Hersh 2015). This effect has also been demonstrated in the famous and controversial laboratory experiments of Stanley Milgram and Philip Zimbardo's Stanford Prison experiment:

By the 1970s, psychologists had done a series of studies establishing the social power of groups. They showed, for example, that groups of strangers could persuade people to believe statements that were obviously false. Psychologists had also found that research participants were often willing to obey authority figures even when doing so violated their personal beliefs. The Yale studies by Stanley Milgram in 1963 demonstrated that a majority of ordinary citizens would continually shock an innocent man, even up to near-lethal levels, if commanded to do so by someone acting as an authority. The "authority" figure in this case was merely a high-school biology teacher who wore a lab coat and acted in an official manner. The majority of people shocked their victims over and over again despite increasingly desperate pleas to stop (Zimbardo 2007b; cf. Zimbardo 2007a, Ch. 11 Milgram 1974 and Bauman 1989, Ch. 6).

The problem with Detective Harris's commands in *Training Day*, as well as the many other examples, seems to be that they are benefiting Harris, the authority, instead of 'serving and protecting' the members of the community. Harris is a tyrant in the sense that he "rules in accordance with his own will and interest" (Arden 2000, p. 467). He is a man become wolf, as Plato might say (Plato 1997c, 565d-566a). The content of his commands lacks justification.

We can also think of the opposite being the case, viz., a seemingly justified command being given by someone without the standing to give such a command. For example, take Mr. Kurtz from Joseph Conrad's *Heart of Darkness*. As we learn from Marlow through the nameless narrator, Kurtz has become a sort of pseudo-god to the native inhabitants of a remote part of central Africa.

Now imagine a 'free thinking' native who does not fall under Kurtz's spell. Kurtz might command him to  $\phi$ , where  $\phi$ -ing is a duty he already has. Perhaps, Kurtz, in an attempt to fulfil his dream of 'civilising' the natives, which is evident in the book he has written, commands the native to develop his musical talents by learning to play classical violin. Kurtz certainly has no standing to demand such things. The native could now respond rightfully to this command with: "Who are you to demand that?"

Assuming that developing one's talents is a genuine duty, the content of the command seemingly applies by giving a concrete direction to the native to discharge his duty and the native indeed has an obligation to do so. However, and here is the point: Kurtz was not justified in issuing the command to the native. Perhaps, the native even has reason to follow the command, but the important point is that the reason he has does not derive from the command itself. In other words, there is no authoritative command for him to conform to.

We thus see that, in two ways, the command itself makes no normative difference. First, as the native already has a duty to develop his talents, the command has not given him a new reason to act. Second, the command does not change the normative relationship he has in accordance with the reasons that exist for him to develop his talents. This second way of making no normative difference is the most important for reasons that will become clear in the next section. The point being that there seems to be an additional condition on

legitimate authority beyond the content being justified. The person issuing the commands needs to (in some way) be in the right position to issue those commands to a particular addressee, i.e., having standing.

To clarify, in the above examples, neither command is meant to be an authoritative command, as they both misfire. My intention was to illustrate that commands can misfire for two distinct reasons. The first one misfires due to the content being unjustified; the other misfires due to lack of standing.

The question of content is relatively straightforward even with disagreement over how to answer it; put differently, we understand what the question is asking. On the other hand, I take it that the question of standing is less straightforward as well as potentially more troubling. As we have seen, someone who actually has standing may command dubious actions and the fact that they have the standing to issue commands imparts an aura of legitimacy on to the particular command. This concern does not arise with the question of content for if the question of content is answered, *ipso facto*, then the command's content is justified. This is entirely right insofar as it is precisely what the question of content is meant to answer. Therefore, a few remarks are in order about what is being asked in terms of the question of standing.

Standing, as rendered here, is a relational concept insofar as commands are always given by the authority to the addressee of that authority. There is no reason why the authority or the addressee of authority needs to be an individual. How an individual can be related to another in an authority relation is quite clear: *A* has authority over *B*. What might be less clear is the cases of a collective. A few words on collectives are therefore in order.

We will assume for the moment that the United States Congress is indeed a legitimate authority for the United States citizenry as a collective. As a collective, the

Congress of the United States issues authoritative commands both to the individual citizens of the United States and to the United States as a whole. An example of the U.S. Congress issuing commands to individuals would be commanding things such as the side of the road one is to drive on or how much one is to pay in taxes.

Another example is the U.S. Congress's ability to declare war. By declaring war, the U.S. Congress changes, through a declarative act, the state in which the United States is in as a collective. This changes the U.S. from a state of peace to a state of war and will have implications for particular citizens to be sure. However, the primary addressee of the declaration is the collective entity, the United States. An enquiry about the standing to issue commands is one that will therefore consider the conditions that allow authorities - e.g., the U.S. Congress - to issue commands to its addressees, so making the addressees accountable to following the authority's commands.

Further, the standing relationship is a *normative* relationship.\* R.B. Friedman once made this point about authority when he said that authority, “implies that there exists some mutually recognized normative relationship giving the one the right to command or speak and the other the duty to obey” (Friedman 1990, p. 71). We can bracket Friedman's particular view. It is not my intention to defend his particular account. The important point for us concerns the question of the normative relationship that is implied by authority, that which can be distinguished from the justification of the content of a command. The normative relationship is that which then enables authorities to give commands in the first place.

In this way, we can see that the question of standing is not what David Schmdtz has

\* In the next chapter we will see how Darwall grounds this normative relationship in the second-person standpoint. In Chapter 5, I will argue that this relationship must be grounded in a more fully developed social ontology.



called 'emergent justification' (Schmidtz 1996, pp. 81-97). For Schmitz, the “emergent approach takes justification to be an emergent property of the process by which institutions arise” (ibid., p. 82). In other words, this approach looks at the pedigree of an institution - for example, in regard to consent (ibid.) - for if an institution is consented to then it is justified. However, this is not the issue of standing, even if it goes some way to confronting it. Rather, as we have seen, the question of standing is a matter of why authorities are entitled to demand anything of their addressees at all. Therefore, it is not the external question of the justification of an institution but the internal question of a particular act of commanding.

The question of standing is therefore: What needs to be the case for someone in authority to be entitled to command their addressees and for the addressees to be obligated to defer? Or, what kind of relationship gives someone authority over someone else? Just as the content of the commands must be justified, so too does the standing required to issue commands. Before proceeding, let us try to be more precise about the type of authority with whom we are concerned and what this type of authority is actually doing when it issues commands.

### 3. Having Authority: Epistemic vs. Practical Authority

What do we mean when we think of someone as being *in* authority? What is someone doing when they are *in* a position of authority? The 'in' is suggestive. It is helpful to distinguish being *in* authority from someone being *an* authority (Friedman 1990 and Guess 2001 Ch 1. §5). *An* authority – for instance, someone with epistemic authority - possesses theoretical knowledge of a specific domain. A physician, to give an example, is a reliable source of knowledge on health. A physician, therefore, is *an* authority on health.

Moreover, an epistemic authority has essentially an advisory or counsel role, capable of delivering hypothetical directives. If we are committed to being healthy, the physician can advise us on how to maintain or regain our health which, in turn, gives us a reason to follow the physician's advice. These reasons are not derived from authoritative commands. Rather, the force of this type of authority turns on linking up with specific commitments which an individual already holds. If the individual changes her commitments then authority's advice fails to have force. However, this is not the structure of when someone *in* authority gives a command.

Amongst others, Stephen Darwall has noted that the distinction between practical authority and epistemic authority is “the point of Hobbes's famous distinction between ‘command’ and ‘counsel’” (Darwall 2006, pp. 12-13). To quote the first paragraph of Chapter 14 of Hobbes's *On the Citizen*:

The distinction between *advice* and *law* is to be sought in the difference between *advice* and *command* [*mandatum*]. ADVICE is an *instruction or precept* [*praeceptum*] in which the reason for following it is drawn *from the matter itself*. But a COMMAND is an *instruction* in which the reason for following it is drawn from *the will of the instructor*. For one can only properly say: *This is what I want, this is my order*, it will stand for reason. But since laws are obeyed not for their content, but because of the will of the instructor, *law* is not *advice* but *command*, and is defined thus: LAW is *a command of that person (whether man or council) whose instruction is the reason for obedience* [...] For *law* comes from one who has power over those whom he instructs, *advice* from one who does not have power. To do what one is instructed by *law* is a matter of *duty*; to take *advice* is *discretionary*. *Advice* is directed to the purpose of *the person instructed*, *law* to the purpose of *the instructor*. *Advice* is addressed only to *those who want it*, *law* also to *those who do not want it*. Finally the right to *give advice* is cancelled at the discretion of *its recipient*; the right of the *lawgiver* is not cancelled at the discretion of the person on whom *law* is imposed (Hobbes 1998, pp. 153-154; cf. Hobbes 1996, Pt. II Ch. XXV).

When someone in a position of authority gives a command then they are giving the addressee a new reason for action to which the addressee ought to conform. The reason is “drawn from *the will of the instructor*” (ibid.); that is, the reason one ought to conform to

the command does not refer to the *content* of the individual commitments one happens to have. One ought to conform to the command because it was the *will* of the authority; one ought to conform to the command because it was commanded. It is the 'say-so' of the authority that provides the reason for action.

We can see an example of this difference playing out in the current debate in the United States about vaccinations for children. There is a growing number of parents who are opting not to have their children vaccinated for measles which, predictably, has caused an increase in the number of cases of the disease. There is overwhelming support for vaccination in the medical community, with the majority of doctors recommending vaccinations and assuring parents that they are both safe for their children as well as necessary to maintain so-called 'herd immunity'.

Here, the recommendation of the medical community is clearly a case of epistemic authority. They are experts on health care issues and, in order to maintain the health of children and prevent cases of measles, we ought to accept their advice. Now, many are calling for local governments to reduce the number of exemptions given to parents. There are calls for parents to be forced to immunise their children – for example, there is a bill currently being considered in California which would, if passed, end the so-called 'personal belief' exemption. These actions are not a case of epistemic authority. If this law passes then the normative landscape is augmented (assuming California is a legitimate authority) and parents ought to defer to the commands of the state, not because of their commitments to keep their children healthy, but because it is the will of the state. What accounts for the difference?

*An* authority is evidently not what we mean when we are thinking about the type of authority that a head of state, the ordinary police officer or a supervisor claims to possess.

For we typically do not think of them as counselling us nor do we take their commands to be hypothetical. What then does it mean to be *in* authority? Being commanded to get your children vaccinated does not imply: 'get your children vaccinated if you want to keep them healthy and belief that vaccinations are safe.' Rather, the command itself is categorical: 'you ought to get your children vaccinated.' Being *in* authority, then, has to do with giving first-order reasons to act that would not exist independently of the command. This new first-order reason to act - the reason given by the command of a legitimate authority - may or may not depend on other first-order reasons to act.

This is a matter that will be addressed when we turn to how authority is to be justified. We are presently focused on what authority is; what is it that needs justifying. Being *in* authority, in the words of David Enoch, then entails the power to 'give reasons robustly', i.e., "the ability to bring about a change in the addressee's reasons for action" (Enoch 2012, p.5; cf. Marmor 2010, pp. 240-241 and Raz 1979, pp. 16-19). Someone *in* authority has the ability to change the normative circumstances of those whom the authority commands; thus, practical authority makes a difference (Raz 1986, pp. 30-31). This is what is typically meant by practical authority.

To clarify what it would mean to 'give reasons robustly' let us turn Joseph Raz's 'pre-emptive thesis'. It states:

The fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them (Raz 1986, p. 46; cf. Raz 2009, pp. 140-142).

This thesis is about what authorities are doing when they issue commands, and how they are changing the normative circumstances of the recipients of their commands. The thesis consists of two main clauses outlined below.

First, an authoritative command gives a first-order reason to act, which replaces other reasons to act (the replacement clause). That is, the addressee ought to act on the authoritative command itself and not on other reasons. If people are asked why they are getting their child vaccinated, the appropriate answer is (can be) '*because* it is the law'. The command itself is the reason for action. This clause is important because it gives addressees who, having no reason prior to the command to act in conformity to the content of the command itself, a reason to do so. These reasons are derived from authoritative commands.

One way to make sense of authoritative commands is to consider what Montaigne says in his essay "On Experience": "[i]f anyone obeys them [commands] only when they are just, then he fails to obey them for just the reason he must!" (Montaigne 2003, p. 1216). Montaigne's formulation in the B edition from 1588 – the last edition published in his lifetime – is even more forceful: "If anyone *obeys the law because it is just, obeys it not.*" (ibid., n39). This is not to say that you cannot act in conformity with the command for other reasons, for example, that the command is just. Authorities cannot tell, just as in general we cannot tell, the particular reason that a person acts upon. Without further investigation, we cannot tell the difference between a parent who vaccinates their child because it is the law and a parent who vaccinates their child because they are following the advice of the medical community. Even with further investigation, it is not obvious that we can ever determine the reason that was actually acted upon. The point is rather that if you were not going to act in the way commanded before the command, or if you were unsure as to how to act, the mere say-so of the command give you a reason to do so. In other words, the command ends further deliberation. This is therefore connected to the second clause.

Second, with the exclusivity clause, an authoritative command also gives an exclusionary reason not to act for a certain range of other reasons (Raz 2009, §III.B; Raz,

1986 pp. 41-53; Herskovitz 2003, §§ 1-2). That is, when a practical authority issues a command, it not only replaces other reasons to act but also excludes (at least some) other reasons from being considered (Raz 1979, pp. 22). They give one “reasons not to act for certain reasons” (Raz 1999, pp. 183). If one is to decide whether to get one’s child vaccinated, one ought not to consider reasons that conflict with the authoritative command, e.g. fear about the health consequences perpetuated by questionable science or the advice of Hollywood film stars. These reasons are not to be weighed against the command to get children vaccinated.

It is the combination of these two clauses that gives authorities the power to change normative circumstances and to give authoritative commands. Practical authorities give one first-order reasons to act which did not exist prior to the command, as well as an exclusionary reason not to consider a certain range of other reasons. There still might be moral reasons why one should consider and weigh against authoritative commands. This is why we no longer find acceptable the so-called ‘Nuremberg defence’ of ‘I was just following orders’ as a moral defence for following immoral commands. We still expect individuals to use judgement when it comes to the moral acceptability of a command even if the command is valid from the point of view of the institution from which it was issued.

If we take the distinction between command and advice together with the pre-emptive thesis, we see that this advice does not entail exclusionary reasons or first-order reasons to act. That is why advice is hypothetical; advice gives you an ability to weigh your reasons to act in a different, more accurate, way. It alerts you to reasons to which you did not have access before the advice was given. This is not the case for a command. A command excludes other types of reasons from consideration as well as robustly giving a reason to act to which one ought to conform. Thus, a command is not hypothetical in the

way that advice is, but rather, categorical.

#### 4. Normative Magic and the Mystical Foundations of Authority

So far, we have looked at what authorities are claiming to do when they give commands. Many are troubled by this extraordinary ability authorities that claim to possess – the ability to change normative circumstances merely by 'say-so'. An ability that legitimate authorities, if there are any, will in fact possess. With this ability it would seem as though authorities were capable of performing a kind of 'normative magic' (Enoch 2012). Authorities are able, through their mere say-so, conjure reasons for their addressees and augment the normative landscape – connecting to the Latin roots of *auctoritas* deriving “from the verb *augere*, “augment”” (Arendt 2000, p.486). By mere ‘say-so’, authorities would create new reasons for action that augment the moral landscape. How then are we to make sense of the idea that authoritative commands create new reasons for action?

Although it would appear as if this were a magical ability, this aspect of commands can be demystified and accounted for in a quite non-mystical manner. David Enoch has developed one way to demystify this ability in his account of 'robust reason-giving'; what we have been calling authoritative commands: “the intention to give a reason *merely by the forming of the intention to give a reason*” (Enoch 2012, p. 7). Let us look at his account of how this magic is possible, i.e., how it is not magical at all. I will then go on to criticise it. One of Enoch's favourite examples of authoritative commands is that of a parent instructing a child to go to bed. According to Enoch, before the parent gave the instruction to the child, there was no reason for the child to go to bed. By the ‘say-so’ of the parent and by an act of will on the part of the parent, the child, suddenly, has a new reason to go to bed – now we can see why we might want to call this normative magic!

Enoch attempts to account for the possibility of authoritative commands in terms of what he calls 'triggering reasons'. A triggering reason, according to Enoch, is when a change in non-normative circumstances changes one's reasons for action. An example of this might be when a bus company decides to change their bus schedules. The company did not intend to give you, the traveller, a reason for action. However, since the bus is now scheduled to arrive ten minutes earlier, and you have a reason to catch that bus, you now have a reason to leave ten minutes earlier. The change in the bus schedule has triggered a reason for why you had to leave the house at a certain time. Notice that one may change one's commitments, i.e., the commitment to take the bus; here the triggering reason no longer applies. However, this is not going to be the case with all triggering reasons. There may be normative background conditions that agents are not free to change.

Although authoritative commands, for Enoch, are a species of triggering reasons, he claims that they are still a “unique particular instance of triggering reason-giving” (Enoch 2011, pp. 2 and §§4.1-4.2). They are distinct in the way in which they relate to the intentions of the reason-giver. What is special about authoritative commands is “the intention to give a reason *merely by the very forming of the intention to give a reason*” (Enoch 2012, pp. 7). Compare this to the bus schedule example above. Opposed to the bus company, which has no intention of giving you a reason, it is the intention of authorities to give you reason. A parent, according to Enoch, who tells her child to go to bed intends to give the child a reason to go to bed and succeeds in doing so when that intention triggers a conditional reason the child already possesses. Enoch gives the following formalised analysis on how authoritative commands are given:

- (i) A intends to give B a reason to  $\phi$ , and A communicates this intention to B;
- (ii) A intends B to recognise this intention;
- (iii) A intends B's given reason to  $\phi$  to depend in an appropriate way on B's



recognition of A's communicated intention to give B a reason to  $\phi$  (Enoch 2011, pp. 7-8; cf. Enoch 2012, §4.4).

Yet, we can already see that Enoch's account is part of a much larger phenomenon.

His words closely resemble those of H.L.A. Hart:

In many different situations in social life one person may express a wish that another person should do or abstain from doing something. When this wish is expressed not merely as a piece of interesting information or deliberate self-revelation but with the intention that the person addressed should conform to the wish, it is customary in English and many other languages, though not necessary, to use a special linguistic form called the *imperative mood*, 'Go home!' 'Come Here!' 'Stop!' 'Do not kill him!' The social situations in which we thus address others in imperative form are extremely diverse... (Hart 2012, pp. 18)

In other words, the issue of authoritative commands, in Enoch's sense, seems to be merely the general phenomenon of declaratives. How is it possible just by the mere intention of uttering something in the imperative mood that can one actually change the normative circumstances of the addressee?

Enoch provides us with two success conditions for authoritative commands. The first success condition, which is non-normative, reads:

“For A's attempt to robustly give B a reason to  $\phi$  to succeed, B must recognize A's above specified intentions, and furthermore B must allow these intentions to play an appropriate role in his practical reasoning” (Enoch 2011, p. 20; cf. Enoch 2012, p. 8).

However, we might wonder if mere recognition of the intention, allowing it to play an appropriate role is enough to do the work Enoch wants. This condition will be the topic of the latter part of the next chapter when we look at how it is possible that subjects are actually addressable to the reasons authorities intend to give.

What we will turn to first is the second condition that Enoch outlines. This second condition is a 'normative' success-condition. It reads:

the attempt must make it the case that a reasons to  $\phi$  really does emerge (in the appropriate way). And we already know that whether this procedure will result in there being a reason to  $\phi$  here will depend on there being an independent reason that

is triggered by this procedure – roughly, a reason (for B) to do as A intends that B have a reason to do (Enoch 2012, pp. 9; cf. Enoch 2011, pp. 20).

There are several things to notice about this 'normative condition'. First, the similarities between it and Raz's 'Normal Justification Thesis' are evident and will be the subject of the first part of the next chapter. Basically, what Enoch is saying is that for a reason to be successfully given it must rely upon the reasons the subject already has. In this way it is unclear why this is a success condition for the giving of a reason. We could accept that it is a condition for giving the reason in a different way; that is, as would be suggested by Raz, it is a condition of justification. What is the difference being made here?

Returning to the case of the bus company changing the bus schedule, the bus company may not be justified in changing the bus schedule. In fact, the current schedule might be the best schedule possible. Yet, they did change it and so you now have the reason to leave earlier. It does not matter if the bus company was justified or not in changing the schedule. What matters is that they did. If you do not follow this new reason because, say, you regarded it as unfair then you will miss the bus. Can't we say the same thing about commands? In order to answer this question, we will need to look more closely at this condition and what it means for the content of a command to be justified based on the reasons one already has in one's possession.

## ***Two: Two Individualist Conceptions of Authority***

This chapter will examine and critique two individualist conceptions of authority. Each conception focuses on only one of the questions facing authority. The first, Joseph Raz's service conception of authority, focuses on the question of content. By looking at this account we will see that any account that focuses on this question alone has trouble answering the question of standing. As we will see, it is hard to understand how such a conception can make sense of the idea of standing, or rather, how it is that authorities can actually give pre-emptive reasons without having the standing to do so.

We will then turn to a conception which focuses primarily on the question of standing. This will be Stephen Darwall's second-person conception of authority. However, Darwall's conception of standing will be shown to be inadequate in answering the question it sets out to answer. That is, although Darwall's conception is meant to answer the question of standing, it fails to do so on its own terms. However, there is still a lesson to be drawn from it and this will be important for the following chapters.

### **1. The Service Conception of Authority: An Answer to the Question of Content**

As we saw in the last chapter, for Enoch's second condition for authoritative commands to be successful an independent reason must be triggered. This amounts to Raz's 'Normal Justification Thesis' (NJT). Let us now look more closely at Raz's account of authority.

Raz starts by helpfully putting forth the following moral question: “how can it ever be that one has a duty to subject one's will and judgement to those of another?” (Raz 2009,

p. 135; cf. Raz 1979 Ch. 1) The centrepiece of Raz's explanation of how this is possible is what he terms the NJT. His justification reads, with a slight modification, as follows:

the normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to [conform] with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly (Raz 1986, p. 53).\*

This is an explanation of 'the ideal exercise of authority' (ibid., p. 42), i.e., Raz's explanation of when authority is legitimate. Concisely put, authority is justified if and only if it enables subjects to act in conformity with what is required by background reasons better than the subjects could on their own. These are background reasons because, as NJT suggests, they are not followed directly; rather, they are pre-empted by the authoritative command which is to be followed. Raz calls his conception the 'service conception of authority', being that its main justification is that it serves to benefit those subject to it: "It is to help them act on reasons which [already] bind them" (ibid., p. 56).

The NJT is one of a duo of mutually reinforcing theses (ibid., p. 55), the second part being what Raz calls the 'dependence thesis':

all authoritative directives should be based on reasons which already independently apply to the subjects of the directives and are relevant to their action in the circumstances covered by the directive. (ibid., p. 47)

We can see how close this is to the NJT. The 'dependence thesis' states what kind of reasons the justification of authority should be based on, viz., the subject's background reasons. A

\* I have, following Scott Herschovitz, as well as Raz's more recent formulation, replaced the original "comply" with "conform" which is more accurate to Raz's original intention (Herschovitz 2003, pp. 206-207). Raz recently reformulated the thesis in "The Problem of Authority", it now reads: "that the subject would better conform to reasons that apply to him any way (that is, to reasons other than the directives of the authority) if he intends to be guided by the authority's directives than if he does not." (Raz 2009, p. 137-138). This is important because Raz makes a technical distinction between "complying" and "conforming" to reasons. Herschovitz nicely sums up the distinction in the following way: a person *conforms* with a reason if she does what the reason requires, whereas a person *complies* with a reason if she acts for it (Herschovitz 2003, p. 202; Raz 1999, pp. 178-179).

command is to be justified by reasons for action that exist independently of the command. Hence, the NJT is an evaluative judgement on how well an authority accomplishes this task. If someone claiming authority successfully gives commands based on background reasons, then, according to NJT, they are a legitimate authority.

In a recent restatement of the service conception of authority, Raz adds a further condition, something he calls 'the independence condition' to his account, reinforcing his overall point. It reads as follows:

that the matters regarding which the first condition [NJT] is met are such that with respect to them it is better to conform to reason than to decide for oneself, unaided by authority. (Raz 2009, p. 137)

This added condition reinforces an important part of the NJT, one intended to avoid a problem which Enoch's account is liable to encounter. Enoch's account would make it impossible for authorities to make mistakes; authoritative commands would only count if they match the reasons that already exist. This is a high price to pay and it is doubtful that any authority could ever live up to such a high standard. It conflates legitimate authority with a completely just authority so the concept of legitimacy becomes redundant. Surely what we are after when it comes to legitimacy is something less than justice; a standard or criterion which authorities should meet even if they are unable to be fully just. It would leave no room for authorities to make mistakes. Raz avoids this complication (and Enoch could adopt a similar strategy) by qualifying his account with 'likely better to conform'; this makes the justification turn on a counter-factual rather than the necessary connection that Enoch seems to favour.

We must be careful with what Raz means by claiming that authorities have to serve their subjects. On Raz's account, the reasons for action are not synonymous with individual interests. His claim that authoritative commands should reflect, and help subjects conform

to reasons they already possess should be clearly distinguished from the similar sounding claim that authoritative commands should advance the interests of their subjects (Raz 1986, p. 48). There are many instances within Raz's account in which an authoritative command helps its subjects conform to reasons they already possess but are also against a subject's individual interests. Raz's example is that of a military commander giving orders against the individual interest of her soldiers. The command may still help the soldiers conform to reasons, viz., the defence of their country, even though it goes against their personal interest and so their survival (ibid.).

One thing that is troubling about NJT and Enoch's success condition is that they rely on the existence of separate, distinct reasons that the subject already has and that the command must match up with. I do not think the right way to account for the triggering aspect is by enumerating conditional duties. That is, Enoch thinks that *A* can only give authoritative commands if and only if *B* already has a specific conditional reason to conform to what *A* commands. In the case of the parent telling a child to go to bed, Enoch says that the child perhaps had the reason all along 'to-go-to-bed-if-you-tell-him-to': "a reason that you successfully triggered precisely by telling him to go to his room" (Enoch 2012, p. 11). I therefore have trouble making sense of this example by Enoch.

This particular example does, for the most part, seem to be a better described case in terms of epistemic authority, insofar as there does not seem to be a reason created. Remember that the difference between epistemic authority and practical authority turns largely on whether a new reason is created. Is it really plausible that, in this case, a new reason is created for the child to go to bed? Is this really the best way to think about the relationship between a parent and a child? I would suggest not. Rather, the child does not (yet) have capacity to properly recognise the reasons that apply. What the ideal parent, I

presume, would do in these cases would be to attempt to explain to the child why she should go to bed (cf. Edmundson 2013, p. 53). Part of the reason for thinking that this is the case is that a parent is not always interested in enforcing compliance. Rather, they are interested in being a model for the child; a source of advice and counsel.

Even in bracketing this worry about epistemic and practical authority, there is still a problem with any account that attempts to justify commands based on specific reasons that an agent already possesses. For a determinate answer to the question of content, such an account relies on a consensus about what reasons exist. Granted, Raz and Enoch will object claiming that consensus is irrelevant; for them it is a matter of fact whether there are these reasons to be triggered or not. However, this seems to have little practical relevance for concerns about the actual legitimacy of authorities because of the deep but reasonable disagreement amongst people about what these reasons are. Indeed, it is not clear how we are to determine which reasons actually exist without a full blown account of normative reality, one that will be highly controversial.

What I will argue for in the following chapters is opposed to Raz and Enoch. What we should be looking for is general reason, rather than specific reasons; that which authorities trigger to justify the content of their command. For one, this is a more parsimonious answer and will allow us to create a critical standpoint internal to a particular authority to judge the content of a command. This internal standpoint will give us greater determinacy in the question of content and our judgements about the legitimacy of particular authorities. However, we will have to wait for this positive story.

Raz's and Enoch's point is that this ability of authorities to create reasons needs justifying. It is not enough for the authority to have the intention to give reason; we must have a clear criterion as to the normative success of reason giving. We want to have a

criterion in regard to which reasons are acceptable for authorities to create and to determine those which reasons are not acceptable. That, in barest outline, is Raz's conception of authority. Let us now turn to what I take to be its shortcoming, viz., Raz answering the question of standing.

## 2. Does the Normal Justification Thesis Deliver Command Based Pre-Emptive Reasons?

It should be obvious in the way that this view is characterised that Raz's conception of authority is concerned almost exclusively with content. Thus, to justify authority we look at the command and its relationship to background reasons. For Raz, it is through the question of content that we will find an answer to the question of standing. However, the claim of this section will be that answering the question of content does not provide enough resources to give a compelling answer to the question of standing.

To make clear how Raz's answer works, we can begin by asking the leading question of 'who has authority over whom and for whose benefit':

*A* has legitimate authority over *B* if and only if *A*'s commands successfully get *B* to conform with reasons *B* already had.

Here we can see that *A*'s standing as a legitimate authority over *B* is determined by *A*'s ability to issue justified content to *B* better than *B* or anyone else is able to do. Jonathan Quong has pointed out that this is the originality of the service conception in attempting to answer the question of "Who has the legitimate authority to decide what I must do?" in terms of "What should I do?" (Quong 2011, p. 119). The question of 'who/whom' is given an answer by showing who can deliver the greatest benefit to those commanded (in terms of conforming to background reasons).

Raz still has more to say on the question of standing. For example, he discusses



what he calls 'jurisdiction' in the Postscript to the Second Edition of *Practical Reason and Norms*, saying that

“[i]n deciding whether one ought to obey the authority's directives, one ought to exclude all the reasons both for and against [ $\phi$ -ing] which were *within the jurisdiction of the authority*” (Raz 1999, p. 192 emphasis added).

It is unclear exactly what Raz means here by 'jurisdiction', but I will take it to refer to the domain in which the authority has standing to issue commands. Yet, Raz says nothing more about how one is justified in having 'jurisdiction' or standing.

However, he does acknowledge the importance of the question of standing in particular cases, viz., when there are two claims of authority that are incompatible (Raz 1986, p. 57). When Raz explicitly discusses the who/whom question, he says it is a matter of evaluating individual cases (Raz 2009, p. 158-159). He claims that whichever authority can do a better job helping subjects conform to their reasons has a “better claim to be recognized” (Raz 1986, p. 57). For Raz, then, it seems that standing, for the most part, is not an issue. Most of the time there are no conflicting claims for authority, and when there are conflicts, an appeal to the NJT will resolve the issue.

Many commentators of late have challenged Raz's answer to the standing question (Darwall, 2013b; Hershovitz 2011; Marmor 2010 and 2011). They object that potentially anyone can be a legitimate authority over others by virtue of being in a position to get their subjects to conform better to reasons they already possess. This potential can be seen most clearly when Raz discusses what he calls the 'scope of authority'. Raz tells us that the scope of authority, “all depends on the person over whom authority is supposed to be exercised: his knowledge, strength of will, his reliability in various aspects of life, and on the government in question” (Raz 1986, p. 73). That is, legitimate authority and the obligation to defer are a matter of a “special relationship between an individual and his state” (ibid., p.

104). However, this so-called 'special relationship' only amounts to a reformulation of the NJT. Is this plausible? Is it plausible to base standing on the justification of content?

Let us focus on one of these challenges, viz. Stephen Darwall. The way Darwall frames his objection to Raz is by asking: is the NJT really enough to give pre-emptive reasons (Darwall 2013a, pp. 146, 160)? Darwall's answer is, of course, the NJT does not give pre-emptive reasons. Remember, it is Raz's view that commands function precisely by giving pre-emptive reasons. If NJT does not show that commands do in fact give pre-emptive reasons then, by Raz's own standard, there is something missing from his account. It seems to me that this is a powerful challenge to Raz's account. Let us see why.

Darwall begins with the three following theses:

- I. *B* will do better in complying with reasons if he treats *A*'s directives as giving him pre-emptive reasons
- II. There is reason for *B* to treat *A*'s directives as giving him pre-emptive reasons
- III. *A*'s directives actually do give *B* pre-emptive reasons. (ibid., p. 160)

Darwall grants, for the sake of argument, I and II but denies that III follows from the truth of I and II. All that I and II (which amounts to the NJT) establish is that *B* has good, independent reasons to follow *A*'s commands. Although it may be foolish for *B* not to  $\phi$ , or not to follow *A*'s command to  $\phi$ , more needs to be said in order to establish that *B* is accountable to *A* for following the command to  $\phi$ . Pragmatic reasons to follow a command do not seem to be enough. The claim is that all Raz can establish is that it is good for the subject to follow the command; not that the command is actually authoritative.

For Darwall, it is a conceptual matter that *B*'s responsibility to *A* (i.e. follow *A*'s commands) is correlated with *A*'s right to hold *B* accountable. It is Darwall's contention that Raz's account does not supply enough resources to understand this normative relationship between the authority and the addressee. Thus, Raz does not sufficiently answer the

question of standing and *A* is not entitled to *B*'s  $\phi$ -ing, even if NJT holds.

This problem comes out most clearly in an example Darwall takes from Raz. In *The Morality of Freedom*, Raz gives an example of John who is an expert in cooking Chinese food. According to Raz's NJT, John would not have authority over Raz unless his goal consisted in “nothing but to prepare the best Chinese meal [Raz] can manage” in which case “[Raz] should just follow John's instructions” (Raz 1986, p. 64). Darwall modifies the example in the following way: Raz, in fact, has “no reasons to do anything other than prepare the best Chinese meal” (Darwall 2013a, p. 147). Therefore Raz would be foolish not to *treat* John's directives as pre-emptive reasons.

But we can ask, with Darwall, does John actually have authority over Raz? In other words, is there really a pre-emptive reason for Raz to defer to John? It seems that there is not. Even though Raz has exclusionary reasons to follow John's commands, they do not seem to have anything to do with John's commands. It is not an exclusionary reason derived from authoritative commands. Rather, we are left with John being an epistemic authority.

If we recall the discussion of pre-emptive reasons, there are two clauses, the 'replacement clause' and the 'exclusionary clause'. What is missing in this situation is the replacement clause: How is it that John can be entitled to Raz's doing anything in the first place? With the replacement clause, authorities are giving new reasons upon which the agent should act. It doesn't seem that John has the standing of a legitimate authority because his commands do not create a *new* reason for Raz. Hence, there is no establishment of pre-emptive reason. Raz only has pragmatic reasons to follow John's commands and to *treat them as* pre-emptive. What this example is meant to show is that without an already established relationship of accountability there is no justification for John to complain, nor is Raz responsible for deferring, however foolish that might be. What Darwall instead

wants to argue for is that:

the reasons with which following a putative authority's directives secures better compliance already themselves assume background accountability relations that are critical to establish the directives' legitimacy or authority (Darwall 2013a, p. 152).

Darwall puts his object in the slogan: "No preemptive reasons without the standing to hold accountable" (ibid., p. 154).

At best, Raz's reply to Darwall seems weak and unconvincing. He says that

[t]he argument is not that one has those derivative reasons stated in III because it would be good to believe in them. It is that it is good to believe in them because they are there (Raz 2010, p. 299).

Even if this is true, it still does not answer Darwall's fundamental question: How is it that a legitimate authority can be entitled to hold its addressees accountable for non-compliance, and why are the addressees answerable to this particular authority? It might be the case that the command should be followed for the reason Raz states, but he still needs to say more about why authorities are justified in holding individuals accountable. For even Raz admits that a fundamental aspect of practical authority is that "it includes an appeal for compliance by the person(s) subject to the authority" (Raz 1986, pp. 25-26). It is unclear, under Raz's account, why any authority would be justified in demanding compliance.

If this is right, and it looks as if it is, then the question of 'who has authority over whom?' seems to be a question that Raz's account does not currently answer. More needs to be said about accountability and the authority's justification for demanding compliance than can be given from the justification of content alone. Or, as Scott Hershovitz put it,

[t]o establish that there is authority in these cases, Raz must explain how it is that an authority acquires the right to bind subjects in virtue of the fact that [the addressees] would be foolish not to treat [the authority] as if it had that right (Hershovitz, 2011, p. 20).

We ought to now turn to Darwall's positive picture of authority since this is the precise

question he sets out to answer.

### 3. The Second-Person Standpoint in General

As we have just seen, Raz's service conception of authority does not adequately answer the question of standing. Darwall has developed a second-person conception of authority, which he claims capable of giving an adequate answer to this question. His fundamental claim about authority is that we need to give an account of who is accountable to whom, i.e., we need to determine who has standing.

Darwall originally developed his conception of the second-person standpoint as an account of the nature of moral obligation. He draws on several historical figures in order to illustrate the importance he sees in the second-person standpoint. The second-person standpoint has proved to be a highly stimulating and active prospect for several areas of philosophy (Lavin 2014).

One of the figures who has been widely associated with this renewed interest in the second-person is Martin Buber. As a member of the dialogical movement in the early part of the 20<sup>th</sup> century in Germany, he helped to make I-Thou relationship famous with the publication of his *Ich und Du* – though the I-Thou relationship can be traced back at least as far as Feuerbach (Feuerbach 2014 §59; cf. Honneth and Joas 1988, Ch. 1). He also had connections with the likes of Jacobi, Hamann, Wilhelm von Humboldt and Fichte (Theunissen 1986, p. 268; cf. Buber 2002a).

Unlike Darwall, Buber was primarily concerned with fundamental questions of ontology, morality (although not primarily with moral obligation) and theology.\* However, Buber does develop a philosophical anthropology of 'encounter' which will help to

\* Michael Theunissen gives a thorough and helpful exposition of Buber's dialogical ontology as well as the broader dialogical movement in Germany in his *Der Andere* partially translated as *The Other*. The translation omits a particularly helpful chapter on Buber's theological commitments.

illuminate some central themes of Darwall's second-person account. These points of illumination are particularly prominent in Buber's important essay from 1938, entitled “What is Man?” – an essay Darwall explicitly draws on in his essay “Being With” (Darwall 2013b, Ch 6) – which focuses on an attempt to overcome individualistic methodology in understanding this fundamental Kantian question\* without lapsing into collectivism (Buber 2002b, p. 237ff). For Buber the fundamental relationship of the 'I-Thou' occurs in-between individuals where each is completely open to the other in an encounter, yet neither is reified as a thing, as an 'it' (Theunissen 1986, pp. 271-290). Buber does not maintain (as Hilary Putnam has reminded us) that the I-Thou relationship is always a good thing, as there can be 'demonic I-Thou' relations (Putnam 2008, p. 62). The example Buber gives here is Napoleon:

The world 'I' remains the shibboleth of humanity. Napoleon spoke it without the power to relate, but he did speak it as the I of an accomplishment. Those who exert themselves to copy this, merely betray the hopelessness of their own self-contradiction (Buber 1996, p. 119).

It remains obscure how exactly to understand Buber here. However, for our purposes, we can leave this complication aside.

The important point for an entry into Darwall's view is that the I-Thou relationship is characterised as a dialogical relationship between particularised and irreducible individuals. In this type of relationship, both individuals are to reciprocate the openness to the particularity of the other: “some degree of mutual access characterises being with or together” (Darwall 2013b, p. 112). This is because, as Buber puts it, “relation is reciprocity [*Gegenseitigkeit*]” (Buber 1996 p. 58).

\* Kant emphasises this as the fundamental question in *The Jäsche Logic* where he adds to the questions from the first critique (“What can I know?”, “What ought I to do?” and “What may I hope?”) the fourth question, “What is man?”, the question of anthropology, which “we could reckon all of this as anthropology, because the first three questions relate to the last one” (Kant 1992, 25; cf. Buber 2002b, pp. 141-142)

To adopt a more Kantian tone, we must see the other as an end-in-itself and not as a mere means to our individual ends; the importance of the other is not their mere use as a thing. By seeing the other as a means to our own ends we turn the other into an 'it', the relation becomes of 'I-it' rather than the 'I-thou': "Every means is an obstacle. Only where all means have disintegrated encounters occur." (Buber 1996 p. 63; cf. Theunissen 1986, p. 275). The thought is that in experiencing the relationship with the other in this way, we must transform the way we live our lives (Putnam 2008, p. 64). We must change the way we live in response to the other.

What we can draw from this for Darwall's account is the following themes. First, like Buber, Darwall is concerned with the encounter between 'You' and 'I' without reducing either to a mere means, an 'It'. Furthermore, this encounter must be reciprocal and mutual. However, this is where the two accounts start to diverge.

In particular, the divergence springs from Darwall's addition of a Rawlsian element to the dialogical picture. Namely, he also understands these particularised, irreducible individuals to be sources of valid claims. Furthermore, the way that Darwall understands the 'I-Thou' relationship is that of free and equal people who encounter each other in dialogue. For example, he puts his view in terms of conversation with reference to the work of Allan Gibbard:

Norm acceptance manifests itself not just in regulating conduct but also in tendencies to avow the norm in contexts of "unconstrained normative discussion." By normative discussion, Gibbard means not just the sort of thing that moralists, moral philosophers, or writers to editorial pages engage in but also something that is virtually ubiquitous in human life, from gossip, to discussion of novels, movies, and sitcoms, to "I was like ...; He was like ..." conversations in which participants display their reactions to others' actions and feelings. In all instances, people negotiate questions of how it makes sense to respond to what people do and what norms for evaluating conduct it makes most sense to accept. And as they do, empathy works to bring others' views inside our perspective so that they can be part of our own critical reflection and not just recorded as what others think. Second-personal

accountability is not the only form of social criticism, of course, but surely much of what human beings discuss concerns what they and we can warrantably expect and demand of one another (Darwall 2006, pp. 170-171 *citation removed*).

It is in this type of dialogue where we find the grounding for ethical relationships. Darwall, amongst others, has taken up this I-Thou, or second-person standpoint, as the ultimate arena in which validity claims are exchanged – it is the space for the game of giving and taking of reasons. Within the second-person standpoint both individuals maintain their autonomy as self-legislating beings whose wills determine themselves, i.e., the Kantian idea of autonomy.

Darwall's conception, then, is based on a dialogical practice of giving and taking of reasons by individuals. Although he does not explicitly posit his account in terms of a dialogical practice of giving and taking of reasons, there are good reasons for thinking that this is indeed what underlies his account – Darwall instead prefers the term 'second-person address'. For example, Darwall opens his *Second-Person Standpoint* by saying that the foundation of morality is “the perspective you and I take up when we make and acknowledge claims on one another's conduct and will” (Darwall 2006, p. 3). Of course this perspective is the second-person perspective. Like the discourse ethics of Habermas (Habermas 1995, Ch. 2 and 1992, Ch. 3) and Karl-Otto Apel (Apel 1996, Ch1), Darwall is interested in working out the pragmatic presuppositions necessarily entailed by this perspective, the perspective of second-person address. He summarises these presuppositions in the following way:

[W]e hold ourselves morally accountable to others when we impose demands on ourselves that we think it sensible to impose on anyone from a perspective that we all can share as free (second-personally competent) and rational. And we presuppose that anyone we hold thus accountable is someone who can in principle also accept and impose these same demands on himself by taking up this impartial second-person perspective and seeing the sense of imposing them on anyone (Darwall 2006, p. 276).



The idea being that what we are investigating when we are investigating the second-person standpoint is the norms entailed in second person discourse. I think we can then say that Darwall is concerned with the practice of the giving and taking of reasons.

Clearly, fundamental to this idea is the Rawlsian conception of persons as “self-originating source of valid claims” (Rawls 1999b, Ch. 16). The main idea behind Darwall's account is that, in order for one person to hold another accountable, the other person must first be able to hold herself accountable. If she cannot hold herself accountable, then attempting to hold her accountable would be a case of 'mere force' and illegitimate. The first person would be disrespecting her as a 'self-originating source of valid claims'.

This brief summary will clearly not do justice to Darwall's full account of moral obligation. However, Darwall does not want his account just to ground ethical relationships and moral obligation. He also thinks that the second-person standpoint can ground the standing which is internal to practical authority (Darwall 2013a; cf. James 2007 and Kar 2011). The following will bracket the moral obligation aspects of Darwall's account. I do not intend to make any commitments as to the adequacy of Darwall's account of moral obligation. The primary focus will be on Darwall's extension of his account to the nature of legitimate practical authority.\*

#### 4. Second-Person Standpoint and Practical Authority: An Answer to the Question of Standing

Darwall's main thesis is that the best way to understand practical authority is through the presupposition that a legitimate authority has the standing to make claims and demands, as well as hold addressees accountable for them. This has led Raz to level the

\* For criticism of Darwall's extending of the second-person see Lavin 2014, §3 and for criticism of Darwall's overall project see “Symposium on Stephen Darwall's *The Second Person Standpoint* in the October 2007 issue of *Ethics* Vol 118, No. 1

charge that Darwall “seems to use “authority” interchangeably with “standing” (Raz 2010, p. 292). Indeed, Darwall does use the phrase “authority or standing” throughout his writings on the second-person standpoint. What Darwall means by this phrase is that practical authority necessarily involves the presupposition of standing, or of a normative background relationship, to make claims and demands as well as “to hold accountable” (2013b p 154). This, Darwall is claiming, is a necessary precondition for the giving of commands. Without this relation, no command would be possible.

The argument Darwall is making is a transcendental one. That is to say, he is trying to answer the question: what are the necessary grounds of the possibility for a successful command?

His answer is that successful commands are necessarily “grounded in (*de jure*) authority relations that an addresser takes to hold between him and his addressee” (Darwall 2006, pp. 3-4). In other words, we must presuppose the relation in order to make sense of commands. For this reason Darwall focuses on the pragmatic structure of the commanding situation, i.e., the second-person standpoint. It will take some time to unpack what exactly this means and what it implies for a theory of practical authority. The rest of this chapter will be devoted to doing so.

Moving, then, directly into our leading question of practical authority: who, according to Darwall, has authority over whom and for whose benefit? The simple answer, which can already be seen from his response to Raz, is:

*A* has authority over *B* if and only if *A* has the standing to command *B* and hold *B* to account for non-compliance.

This captures what Darwall calls 'second-personal address' which is characterised by an irreducible family of concepts: practical authority, responsibility to, valid claim or demand,

second-personal reasons (Darwall 2013a pp. 141 and 151; cf. Darwall 2006, pp. 11-15). Furthermore, 'second-personal address' is irreducibly relational as it is always made by an addresser and "necessarily always *to* someone (an addressee)" (Darwall 2013a, p. 151; cf. Darwall 2006, pp. 11). We may worry that the account relies on 'practical authority' as a fundamental element; in other words, that it cannot then give an account for the concept being presupposed. However, this is not the case. Rather, through elucidating this family of concepts in conjunction we can learn a great deal about how to understand practical authority, particularly about the question of standing.

We can already see how radically different this starting point is from Raz's service conception. While Raz starts with, and focuses primarily on, how to justify the content of authoritative commands, Darwall starts by asking what characterises the background normative relationship which makes practical authority possible. This is clearest in *Second-Person Standpoint* when Darwall, in a footnote, asks:

But what gives the "law" we committed ourselves to normative force? The fact that we committed ourselves to it, as if adopting it together? That could be so *only if there exists a further background normative relation that gave us the authority so to bind ourselves voluntarily* and whose authority does not itself depend on a voluntary commitment (Darwall 2006, pp. 264n26 emphasis added; cf. Ware 2009).

In other words, Darwall's account focuses primarily on the question of standing.

The best way to get a handle on Darwall's second person standpoint and his conception of standing is to start with an example he borrows from Hume when introducing his theory. It is Hume's 'gouty toe' example (Hume 1975, pp. 226): you stepping on my foot. Darwall concedes that there might be several reasons for you to remove your foot from mine (Darwall 2013a, pp. 136-137, 155).

For example, it might be the case that you think suffering is a bad state of affairs to which you do not want to contribute. Based on this fact alone, you might think that you

should not step on feet, as stepping on feet causes undue suffering. Therefore, when you are alerted to the fact that the weight of your foot on mine causes suffering (the suffering here should be phrased third personally because it is irrelevant whose suffering it is), then you find yourself in the best position to alleviate the suffering, and you remove your foot. Alternatively, you might think stepping on feet is “base thing and so beneath” the virtuous person, you aspire to be a virtuous person; therefore, naturally, you move your foot when alerted (Darwall 2010, pp. 262).

According to Darwall, both of these reasons can be accepted “without holding that anyone has any claim to [you] not stepping on other's feet or that it would wrong anyone for [you] to do so” (ibid.). The problem in both of these cases is that there is no-one in a position to give a pre-emptive reason or to hold anyone accountable for non-compliance. In fact, it is not necessary for there to be other people at all, only that there are other beings who can suffer (or have feet). Hence, there is no authority and the reason why there is no authority, and why there cannot possibly be authority in these situations is, according to Darwall, a lack of the normative second-personal relationship that holds between you and I. In other words, I must be entitled to make demands on you and hold you accountable for wronging me, and you must be answerable to my demands – this is just what it means, for Darwall, to be in a second-personal relationship.

Darwall grounds this prior normative relationship in the idea of equality and a common basis of authority. This is what he terms 'Fichte's point', which posits that

any second-personal claim or 'summons' (*Aufforderung*) presupposes a common competence, authority, and, therefore, responsibility as free and rational; a mutual second-personality that addresser and addressee share and that is appropriately recognized reciprocally (Darwall 2006, pp. 21).\*

\* We can leave aside the historical accuracy of Darwall's reading of Fichte's point as he can make this point independently of its purported historical origins but doubts have been raised about the fidelity of Darwall's reading (Ware 2009, pp. 262-282).

'Fichte's point' is supposed to show that for (legitimate) authoritative relationships to arise it is necessary that individuals recognise the other's common dignity. In Rawls's words, that each is a "self-originating source of valid claims" (ibid., Ch 10; Rawls 1999b, Ch. 16).

How does this connect to the gouty toe example? In Darwall's essay on "Fichte and the Second-Person Standpoint", he tells us that:

Since second-personal reasons concern not the goodness or badness of states of the world considered independently of our relation to them, but rather agent's *relations* to one another, they are invariably agent-relative in some way or other. (Darwall 2013b, p. 227)

In other words, what matters for practical authoritative relationships is not the reasons that might already apply to the individuals. What matters is rather the reasons that we justify to each other as free and equal persons.

It is only in you recognising my claim on you that you are made responsible. For you to be responsible for removing your foot from mine, you must recognise as valid my claim upon you; this is also what allows me to hold you accountable. Now we might wonder why we need to think of ourselves as responsible to the other's claim in order to be accountable to him? In other words, if another person makes a valid claim on us, why do we need to accept such a claim in order to be accountable to him for it? However, it is still open to Darwall to say that we are responsible for the content of the claim but this is not his question. Rather, the question he is raising is about who we are accountable to and who can hold us accountable.

Darwall answers this question by defending a further point, viz., 'Pufendorf's point'. This point says that in order to hold someone responsible for an action, that person must already be able to hold herself accountable. In Darwall's words:

owing to the [conceptual] connections that Pufendorf insists on between obligation

and accountability, someone can be under a moral obligation to do something only if he can hold himself to the relevant demand through recognizing its legitimacy [...] Someone can be accountable only by *holding himself* accountable (Darwall 2013b, p. 213; cf. Darwall 2006, pp. 23, 250).

Again, with the gouty toe, you would be responsible for moving your foot when I demand you to do so, only if it is something for which you can already hold yourself accountable.

This point seems, like 'Fichte's point', to be based in a particular reading of Rawls' fundamental claim that people are 'self-originating sources of valid claims'. If I demand that you  $\phi$ , then you have to recognise my demand as something for which you can hold yourself accountable in order for me to hold you accountable for your  $\phi$ -ing. Why? If you do not recognise my demand as something for which you can hold yourself accountable, then I would be violating you as a source of a valid claim. I would be disrespecting you while demanding that you respect me as a source of valid claims, and this would be violating our equality as moral beings for me to treat you as a mere means.

## 5. Coercion, Accountability and the Right to Rule

The relationship between recognition of the others claim as valid and being accountable for that claim becomes clearer if we turn to another claim of Pufendorf's which Darwall also accepts and insists upon.

Darwall insists that there is a conceptual distinction between mere force or coercion and holding someone accountable; a distinction any act of legitimate authority presupposes (Darwall 2006, pp. 250 and Darwall 2013b, pp. 189, 216, 252; cf. Arendt 2000, pp. 463). This distinction amounts to the following: When  $A$  has the standing to demand  $B$  to  $\phi$  and  $B$  fails to comply, it would be legitimate for  $A$  to hold  $B$  to account. Where  $A$  does not have standing to hold  $B$  to account,  $A$ 's commands would become mere coercion "and this would violate the addressee's authority" (Darwall 2006, pp. 250). The reason for this is that  $A$ 's

command cannot be justified to *B* and *B* can only be accountable for a command if and only if *B* can hold *B* accountable. Thus, *B* is not responsible to *A*. *B* has no second-personal reason to follow *A*'s command. Legitimate authority, then, fully turns on presupposing a normative, second-personal, relationship of *A* and *B* – of 'you and I'. *A* can legitimately command *B* if and only if *A* can justify his commands to *B* in the sense that *B* can take himself as responsible to *A*'s demands.

The distinction between accountability and mere force is hugely important for the understanding of authority. In particular, this distinction opens up the possibility of seeing how authorities could legitimately enforce their commands. This is one of the most important elements that Darwall's account helps to highlight. It is particularly important being that the enforceability of commands is one of the core aspects of authority. As John Dunn writes:

what they have to be able to do and go on doing if they are to exist at all, and hence to be in a position to do anything else. The capacity they must create and maintain is the capacity to make each of their subjects in the end act as they direct and not as the subject in question would otherwise be inclined (Dunn 2000, pp. 117-118).

In other words, the authorities not only purport to determine their addressees' commitments but they are also concerned with ensuring that their subjects do, in fact, act as commanded. This is particularly true in the case of our political lives, which is the target of the quote from Dunn. States claim the right to subject their citizens to a certain range of commands and obligations. Yet perhaps, even more problematically, these commands are backed by the threat of physical coercion for non-compliance.

However, the claim that only states have coercive authority should be rejected. It seems that most authorities have coercive means at their disposal to criticise disobedient members and enforce their commands (Lagerspetz 1995, Ch. 4.1). Think of the ability of

businesses to terminate employees and religious institutions in order to excommunicate their members, e.g., the famous *herem* of Spinoza. One might claim that we are free to leave jobs or religious associations much easier than we are able to leave a state. In some cases this may be true, but it does not seem to be true in all cases. In any society, without strong labour rights and a social safety net the poor will not be able to walk away from a job so easily, much as the poor will not be able to leave a state as easily as their wealthy counterparts.

The type of coercion that does seem to be unique to states is, as Weber rightly pointed out, the monopoly on *physical* coercion (Weber 1968, p. 54 and Weber 1946, p. 78). In other words, all authorities claim for themselves the ability to enforce their commands. This is not just something that authorities claim. It is a feature internal to commands themselves. As we have already seen, commands are not suggestions that one takes up as a further consideration in practical reason; instead, they demand conformity by their very nature.

Now, the claim is that if one decides not to conform, a legitimate authority is entitled to extract conformity due to the enforceability which is also internal to the nature of commands (Ripstein 2004; cf. Derrida 1989, pp. 925). The coercive violence of enforceability mediates between the autonomy of the individual and the authority of a command. Kant makes the same point - law and freedom without force is anarchy:

[I]n order for law to be effective and not an empty recommendation, a middle term must be added; namely, force [*Gewalt*], which, when connected with freedom, secures success for these principles (Kant 2007, 7:330).

If a command is not backed up with some sort of sanction for non-compliance then it would seem not to be a command at all, but rather a mere suggestion; one that can be accepted or rejected at will being that this is not how commands are given (that is, commands are not



mere suggestions).

We should be clear about what such an enforcement entails; it is not a further incentive to act. However, the incentive view of enforceability and coercion is common. When one takes up this incentives view, then coercion becomes secondary. This is expressed by Raz, for example, when he says that:

we can imagine other rational beings who may be subject to law, who have, and who would acknowledge that they have, more than enough reasons to obey the law regardless of sanctions (Raz 1999, p. 159; cf. Green 1988, pp. 72-73).

The idea being here that the coercion and the enforceability of commands is only secondary when it comes to understanding authority. In relation to this view, this aspect of commands is to provide extra motivation for addressees to act. The extra motivation is only necessary because we are limited human beings. If we were fully rational then there would be not be a need for this aspect of commands. Hence, this is not the best way to understand coercion and enforceability and how it is related to commands.

Instead, the idea is that the one who is commanded is doing something wrong when they do not conform to the command, while the authority, when legitimate, is entitled to extract conformity through coercion. This is what Darwall helps us to see with the distinction between 'mere force' and justified coercion turning on the addressee of the command becoming accountable to the authority for the conformity.

When the enforcement of a command turn out to be legitimate, it is because the addressee is in fact accountable to the authority to defer. As in Roman law, a creditor has the right to extract payment, with force, from a debtor because the debtor is accountable to the creditor; conversely, the creditor is entitled to what the debtor owes (Ripstein 2004, pp. 7-8). It is this owing relationship which ensures that addressees are in fact accountable for following commands. This is how we should understand the right to rule. It is a directed

relationship where the addressee is accountable to the authority. When this accountability relationship exists, then the authority has the right to extract authority from addressees when the addressees fail to conform.

David Enoch has developed a 'debunking' case against the idea of a right to rule. Enoch objects to the notion of the right to rule generally. In particular, he objects to the idea of the directionality of authority; that is, obedience to the command is owed to the authority (Enoch 2012, §7). If Enoch is right, then Darwall must be wrong. For Darwall, it is the directionality that is necessary to make sense of authority: for you to have authority over me, you must be entitled to make demands on me and for you to be so entitled, I must be accountable to you for those demands. This is the irreducible directionality in Darwall's account. Therefore, it is worth spending some time on this issue.

If we recall Enoch's normative success conditions for authoritative commands, the reasons why Enoch thinks that he does not need the right to rule become clear. For authoritative commands to be successful, it needs to trigger independent reasons that one must already possess. He then claims that the reasons triggered by authorities are not a general reason why one owes the authority obedience. Instead, it is the reasons triggered themselves for which individuals are accountable, explaining why then these reasons will always be conditional reasons. Remember Enoch's explanation of how a parent can successfully give a child reason to go to bed. He claims that the parent is only (normatively) successful if the child has a conditional reason to 'to-go-to-bed-if-you-tell-him-to'. Leaving aside the adequacy of this particular example, we can see why Enoch wants to deny the directionality of authority. It is because, according to Enoch, that for which the addressees are accountable is the reason triggered.

There is no need for the directionality of authority because the individuals whose

reasons have been triggered already have sufficient normative reasons. The authority is there simply as a secondary means to help motivate individuals to conform. Why do they need any further reasons?

However, his reasoning seems weak. For one, most of the force of his claim turns on one example, that of arbitration. Enoch asks, if *A* is arbitrating between *B* and *C* and *A* rules in favour of *B*, whom does *C* owe the duty of obedience to (ibid., p. 29)? Enoch's answer is the same as in the case of a parent triggering a child's conditional reason to go to bed. All *A* does is trigger *C*'s reason for  $\phi$ -ing. Hence, *C* is accountable for  $\phi$ -ing, not to *A* for  $\phi$ -ing.

Notice that arbitration is an institution insofar as it is defined by particular rules about the roles, penalties and moves that give a structure to the activity of arbitration. Obviously what will not make the penalties follow the decision of the arbitrator will vary, based on the larger institutions in which the arbitration is embedded. However, one thing that seems clear is the roles that any arbitrating institution will entail and the general moves that will be allowed. Arbitration is characterised by an arbitrator who has authority over the decision and parties who will give testimony in regard to which the arbitrator will decide between. The question that Enoch is asking is: what does the party in this institutions owe and to whom? His answer is that the two parties, and not the arbitrator, are the only two who could own anything to the other. The arbitrator is not owed anything. No-one is accountable to the arbitrator, to *A*.

Yet, it is not entirely clear why it is that *C* is not accountable to *A*. It is intuitive that *C* owes *B* whatever it is they went to arbitration to determine. One thing that matters to *C* and *B*, then, is determinacy. They both want concrete closure on a deliberative question. Both *B* and *C* are making claims upon the other but disagree as to the validity of these

claims being that they are appealing to a procedure to give a determinate answer.

Now let us take it for granted that both *B* and *C* are acting in good faith; that is, neither is trying to manipulate the other. They both genuinely believe that their own position is justified and the other is mistaken. They need a determinate answer to their question but neither can give such an answer – this is precisely the problem. If they are arguing about whether *C* ought to  $\phi$  or not, where *B* claims that *C* must  $\phi$  and *B* claims that he does not, after *A*'s decision *C* in fact now owes it to *B* to  $\phi$ . This is particularly clear if it is a matter of payment. If *B* claims that *C* owes her \$100 and *A* rules in favour of *B*, then *C* owes *B* \$100.

Presumably, if all they were after was determinacy then they would agree to flip a coin, or some other procedure to give them some determinacy. Yet, due to the fact that they both think they are in the right, they want to make their cases to an objective third party who will be able to give a determinate and intelligent judgement to settle the disagreement.

However, Enoch would argue that this verges on decision fetishism (Enoch 2009, §2). That is, he believes that to focus only on the procedure and how a decision is made is only one thing to be valued in such a decision, and what really matters most is the instrumental value of such procedures.

Further, *B* and *C* both want assurance that the decision, whatever it is, holds. They want the dispute to be over. Neither would bother going to arbitration if they believed that the other would not be accountable for conforming to whatever decision the arbitrator made. This would re-establish the problem which they originally wanted over-come, and they would want the decision to be enforceable.

For Enoch, this cannot actually be enough because, for *A*'s decision to be normatively successful, one of two things must be the case. Either *A* must in fact be *right*

about  $C$  owing  $B$ , or  $C$  (and  $B$ ) must have a conditional reason triggered by  $A$ 's decision 'to- $\phi$ -when- $A$ -decides-against-you'.

However, this latter option is foreclosed by Enoch due to his eschewal of a general right to rule. Why? To have such a reason, it seems that this must be directed or it would otherwise be difficult to make sense of the content of such a reason. With other reasons triggered, we can see why it is not owed to the authority. For example, if we return to the case of the parent 'commanding' a child to go to bed, we see that the content of the triggered reason 'to-go-to-bed-if-you-tell-him-to' is derived from the benefit of getting enough sleep. Alternatively, it could be expressed in the case of arbitration, if the reason given is that of 'you-must-pay-what-you-owe' seems reasonably to derive its content from already owing something. This is the first option.

The first option, viz.,  $A$  must make the right decision, must be Enoch's view. This seems to be too high a standard to place on authorities. Part of the role of arbitrators (and practical authorities more generally) is to make determinate judgements about issues of disagreement. If authorities are not give enough room for error, then their roles would seem to be completely undermined.

Putting aside this issue, there is arguably a deeper issue inherent to Enoch's account. Enoch does not seem to capture fully the phenomenon of the arbitrator situation. This is not surprising as his account is, by design, a debunking account.

What Enoch's account misses is that  $C$  also owes it to  $A$ , as arbitrator, to defer  $A$ 's ruling. In fact both  $B$  and  $C$  owe this to  $A$ , to affirm exactly what Enoch wants to rule out; that is, a general right to rule or to say that both  $B$  and  $C$  have a general reason 'to- $\phi$ -when- $A$ -decides'.

The reason why this must entail the right to rule is that the content of the reason

seems to be that one owes it to *A* to follow what *A* says. The thought would be that the structure of the arbitrating situation is one where the parties make themselves accountable to the arbitrator and that this is constitutive of arbitration itself. Without this element of accountability, it would be difficult to make sense of the idea of arbitration at all.

One way to view this situation is to think about the enforceability of *A*'s decision. In other words, who can compel *C* to make payments? It would seem that the answer is *A*. Is it not the case that, by going to the arbitrator, both *B* and *C* are giving up some of their rights and responsibilities to the arbitrator? By investing *A* with authority in this circumstance, not only do *B* and *C* give up their judgement to *A*, but *A* is also granted the right to enforce that decision.

If the decision was enforceable by *B*, would *B* and *C* not be back in the 'state of nature'? Back into the situation they were trying to escape by going to arbitration? It appears to me, contra Enoch, that there is directionality of obligation here and it does play a large role in aiding our understanding of authority.

Otherwise, it is difficult to understand what force agreeing to arbitration has for either party. Furthermore, this makes much more sense of how to account for the reason being triggered in the case of arbitration. As opposed to holding the arbitrator to some exceedingly high epistemological standard requiring the right answer for the decision to be legitimate (assuming there is even a right answer to all cases of arbitration, which it is not necessarily clear), we are left with the procedure of the arbitration which triggers the 'to-φ-when-*A*-decides-against-you' reason for action, that which is both grounded in *B* and *C*'s agreement to the procedure of arbitration.

What is at issue here is a matter of when an authority actually has the right to rule. Darwall's suggestion seems to amount to the claim that authorities have the right to rule

only when addressees can hold themselves accountable for the commands of the authority.

## 6. Radicalising the Question of Standing: A Critique of Darwall's Second Person Standpoint

For all that Darwall has to say about standing and its necessity for understanding authority relations, it is unclear what ties *A* and *B* together in such a way that *A* has the right to hold *B* accountable in non-moral cases. It is a question of institutions and the type of relation individuals have to them. The question is not what makes it appropriate for *A* to hold *B* morally accountable? Darwall supplies an answer to this question., The question is rather, why should we think that *A* has any right to demand anything from *B* in the first place? Why are *A* and *B* committed to a particular institution?

The answer is clear in Darwall's fundamental moral example of the second-person standpoint, Hume's gouty toe. In stepping on *A*'s foot, *B* seems to have, potentially, already infringed upon *A*. There is already a connection between the two which makes it clear why *A*'s 'summons' - why *A*'s demands - connects to *B*. *B* is already morally interacting with *A*; there is already a reason for *B* to be moved by *A*'s demands. After all, it is *A*'s toe. Thus, through some moral interaction already occurring between *A* and *B*, they are already in the second-person standpoint.

What is unclear is why this type of case can be generalised to other non-moral interactions. What is particularly unclear is why individuals are tied into the institutions where we find practical authorities embedded. That is, why are there bonds between employer and employee, between sergeant and a private or between a state and a citizen? What Darwall fails to deliver in his second-person conception of authority is a story as to why individuals can demand reasons and hold others accountable without there being some

prior moral event or relationship which ties the addressee and addresser together. What Darwall has given us is a story that characterises the background relationship that gives us the authority to bind ourselves. Indeed, this may be the bedrock of moral obligation. How do we move from this potential insight to an account which captures practical authority?

Darwall would claim rightly that what he is concerned about is giving an account of a second-person relationship as such. This may be fine for the case of morality since it seems that moral address can reasonably be assumed to be a universal practice. We do not seem to be able to escape this form of address and from this sort of practice. We are always eligible to be tied up into second-person address by another when the claims are about what is morally unjustifiable. The thought would go, that because we are always already part of the moral community, all our actions must be justifiable to others who are affected by them (cf, Korsgaard 1996a, Ch. 10, Habermas 1992, Ch. 3). If one of our actions is potentially unjustifiable to others who are affected, then we would be disrespecting them as an autonomous person.

For example, if *B* is trying to violate the integrity of *C*, *A* would be justified in demanding *B* to stop because he would (potentially) be doing something immoral – even if *A* must be making this demand on behalf of *C*. *A* would be hindering a hindrance (Kant 1996c, 6:231), However, *B* could reasonable respond that, in fact, *C* agreed or consented to the behaviour or that they were practising a scene in a play or some other excuse. If *A* accepts this reason, then it would look as if *B* was not doing something wrong. If *A* rejects this reason (perhaps *B* is doing something that *C* could not reasonably consent to), then *B* looks accountable and should stop.

It may then be that for purposes of moral obligation, my rights or freedom end where another person begins – and perhaps by extension, where the other person's property



begins. In short, we may need a prior, potentially wrongful, event to occur in order to hold the other responsible. Otherwise, we seem to be left again with the question of standing, a question which Darwall's account is meant to make determinate. We can always ask when someone is making a demand on us, who they are to make such demands?

If this is right, then we would seem to be left with a problem in the case of practical authority; the question of the “would-be independent, who would genuinely prefer life without the state” (Wolff 1995, p. 95). Hence, if we are going to maintain the idea of a person as a 'self-originating source of valid claims', it seems difficult to understand how it is possible for practical authority not to be a case of mere force and hence illegitimate. Another way of putting the point is in terms of the particularity requirement: Why is *B* bound in any way to *A*? What Darwall will need to show is the rational impossibility of a 'would-be independent', otherwise any case of authority is merely contingent and could lapse into mere force. Kant, for example, attempts to give this type of argument in the opening pages of the 'Doctrine of Right'. However, when Darwall discusses individuals in a second-person reason giving situation, he always takes them to accept already the structure which they are in:

Consider, for example, an order delivered by a superior to an inferior within a military chain of command. If a sergeant orders a private to do ten pushups, she addresses a reason to him that presupposes her authority to give the order and the private's obligation to obey it. So far, the only relevant normative presupposition is of unequal authority; the sergeant has the standing to give orders to the private, whereas the private has no standing to give orders to the sergeant. (Darwall 2006, p. 259)

What is clear from the way Darwall sets-up his conception of authority, it must be the case that the private is the one that binds herself and accept the sergeants commands as a free and equal person (ibid, p. 260). For Darwall, the question arising at this point is why the private cannot rescind her commitment to being a private, to being part of the military chain

of command? Why does she not have this right as a free and equal person? If she can bind herself, why should she not be allowed to unbind herself? If this can happen, it would seem as if the sergeant has no grounds to complain. If the sergeant tries at this point to force the private to recognise his role as private, then the sergeant is disrespecting the private as a free and equal person. In other words, the sergeant has lost her authoritative standing over the private and any commands the sergeant now attempts to give and enforce will be cases of mere force and not legitimate authority. Granted, it may be costly but it is unclear that it is either irrational or violates any sort of obligation which Darwall is entitled to assert.

There seem to be three ways in which Darwall can escape this problem. First, he can appeal to an idealising account of what a 'reasonable' person would believe and desire. This 'reasonable person' must exclude the reasonableness of the 'would-be independent'. However, it is unclear how this will work without the classic problems and objections to idealisation coming about (e.g. Gaus pp. 17-18; 318-319 and Quong 2011). Even if these problems can be overcome and the idealisation restriction justifiably excludes many individuals from being considered reasonable, it is hard to see what is conceptually problematic with a reasonable person wanting to be an independent.

Further, this idealisation seems to be in tension with the idea that fundamental claim about individuals being a 'self-originating source of valid claims'. Would it not be the case that we would be stripping a person of their personhood if we rejected his claim out of hand as unreasonable? This is especially true of the case of the would-be independent. What harm in terms of repressing the views of others (Rawls 2005, pp. 61) would the independent be doing to others?

One way to do this without violating the personhood condition, that which seems open to Darwall, would be to put consistency conditions on individuals' beliefs and desires.

This would at first seem a harmless form of idealisation. However, a problem remains. This form of idealisation would still violate the personhood condition because there is always a choice when it comes to resolving inconsistencies. If two beliefs are inconsistent, to regain consistency one only needs to relinquish one's belief. Or in the case of an inconsistency between a belief and an intention, the individual can relinquish either the belief or the intention (Broome 2007). The question of which one of these is relinquished is up to the person. Therefore, with idealisation, even in this minimum form, we risk violating the personhood condition and turning authority into mere force.

Darwall's second option would be to appeal to higher order procedures. This is the strategy taken, for example, in the later Rawls. This higher order strategy, is compelling but fails to deal with the problem of particularity. All of the challenges of the first strategy come back in. How can some form of idealisation fit with the personhood requirement? Why should we think that this particular individual is answerable to these procedures? Why should any person be held to account if they no longer find these particular procedures beneficial or desirable or acceptable basis to cooperate?

Third, and finally, there is the option of traditional consent theory. Even given all its problems, this might go some way to answering why *A* and *B* are bound together. If they are bound together in this way, then it looks like some higher order procedure to determine the content of that relationship could do a lot of the work required. However, once again, the personhood requirement becomes an issue. If it is an individual commitment to join, then why should the individual not maintain the right to rescind that individual commitment? After all, this seems to be something we can do with our normal commitments (more on this in the next chapter).

These are not meant to be insurmountable criticisms of Darwall, but are intended to

raise doubts about his ability to answer the question of standing with the conceptual tools internal to the second-person standpoint.

Let us close with an example which will highlight these problems. We can look at the important issue of climate change. We can agree that climate change is anthropogenic and will cause massive harm if we do not act collectively. There seems to be at least two reasonable strategies about how to deal with such harm. First, there is a market-based strategy of carbon trading. Second, there is a non-market based strategy. One of the major differences between these two strategies has been characterised by Michael Sandel through the moral distinction between fines and fees (Sandel 2013, Ch. 2 and Sandel 2005, Ch. 14). The distinction is that “[f]ines register moral disapproval, whereas fees are prices that imply no moral judgement. When we impose a fine for littering, we’re saying that littering is wrong” (Sandel 2013, pp. 65-66). Many, like Sandel, think reasonably that using more carbon than one ought to (how much carbon one has a right to use is another question), should carry a 'moral stigma'. Those who support a carbon trading policy do not see this type of argument as persuasive (Caney and Hepburn 2011). Neither strategy seems to be unreasonable. Hence, if we were to accept the second-person conception of practical authority, no authority can legitimately arbitrate this issue. If they do, it will be mere coercion. I think this will hold true for any issue in regard to which there is any reasonable disagreement.

Focusing on higher order procedures of the decision makers will not necessarily help. If everyone agreed on the procedures through which the state makes its decisions, then those accept the procedures, like Sandel, must also accept (for the time being) the outcome. Now this might be true for a while. However, consider what would happen if Sandel and those who agreed with him consistently lost out on all decisions or, for some

other reason, began to doubt the validity of the decisions procedure. At a certain point though, according to Darwall's view, it might be reasonable for Sandel to stop supporting the decision procedure and any decision being taken through that procedure would now be a case of mere coercion. In order to stop this from occurring, Darwall would need to give an argument justifying why people like Sandel would not be entitled to change their mind about the acceptability of a decision procedure, particularly in view of the fact that the number of logically possible decision procedures open to any group is "a number that on some estimates exceeds the total number of elementary particles in the universe" (List and Pettit 2011, pp. 49).

The conclusion one could draw is that the bond between the authority and the addressees is made to appear like a pretty weak relationship. So weak, in fact, that it looks like the addressees could renounce their bond to the authority for just about any reason they find valid. The authority would have to accept these reasons, otherwise they would become illegitimate. Darwall's account would be extraordinarily revisionist in regard to how we normally think about the relationship between authorities and their addressees; it is in danger of losing the phenomenon of authority relations as such. What we find then, is that one of the issues that Darwall's conception of authority is meant to address - viz., who is accountable to whom, the question of standing - appears not to be too very promising. Again, this is not to say that this strategy will not work in the case of moral theory. The important point for us is that it does not seem promising to answer this question of standing without some more theoretical work regarding why *A* and *B* are bound together in some way that respects their autonomy as individuals.

In the next chapter, I will give a general argument that will reinforce these doubts and show why any account of standing remaining inside an individualist framework (even

if there is room for a modest sociality like Darwall's) will always face the dual problems of particularity and the independent, requiring that they be overcome.

### ***Three: Authority as Social Practice***

We ended the last chapter by raising some doubts about Darwall's second-person conception of authority and its ability to answer the question of standing, the very question the account is intended to help us answer. Those doubts revolved around why individual people, thought of as self-originating sources of valid claims, are bound to the institution in which authorities are embedded; why are citizens bound to their particular state or employees to their employers?

This was highlighted by the case of the private and the sergeant. It is only in the institution of the military that we can make sense of why the sergeant has the authority to command the private to do push-ups. Now the question is, why is it the case that the person who inhabits the role of private is bound to following the commands of the sergeant? Why should the private not be able to renounce such a role and walk away without the sergeant having any normative reasons to complain?

Darwall does not directly engage with this question. However, there is a growing wealth of writing on exactly this issue. In this chapter, we will look at how individualistic accounts of sociality do not provide enough resources for understanding the binding force of institutions. By way of introduction, we will look at one of the founders of sociology, Max Weber, whose individualistic account of social action has had a lasting impact. Then we will move to the contemporary work on Michael Bratman who shares many methodological assumptions with Weber. What we will see is that these individualistic assumptions make it difficult for these views to properly account for institutions and their ability to bind individuals.

## 1. Modest Sociality and Social Practices: The Inadequacy of Sociality Based on Content

Max Weber's account of social action can be seen as a paradigmatic example of individualistic methodology in understanding human sociality. Weber sees action to be about the individual's subjective meaning, or intention, attached to behaviour (Weber 1978, p. 4; cf Schutz 1967, Gilbert 1992, Ch. II and Greenwood 2003, p. 100). Action is social when it takes into account the behaviour of others (ibid.). From this, Weber defines social relationships in the following way:

The term 'social relationship' will be used to denote behavior of a plurality of actions insofar as, in its meaningful content, the action of each takes account of that of the others and is oriented in these terms (ibid, p. 26).

One way to make sense of Weber's understanding of 'subjective meaning' is in terms of commitments and intentions of the individual (Schutz 1967, p. 6). Weber maintains that we can give a full account of sociality wholly through the means of individual intentions and their 'taking account' of and their 'orientation towards' the behaviour of others.

Weber's definition can indeed be useful in distinguishing between cases of non-social and social actions without asserting that any action involving others is by definition social. In fact, there are several actions involving others, for Weber, which are not considered social action proper. First, there are cases where an individual sees others not as actors, as a 'you', but as an inanimate object, as an 'it'. An example of this, for Weber, might be when a cyclist navigates through a crowd seeing people as things to be avoided. If there is a collision, it can be seen as a "natural event" (Weber 1978, p. 23). A second, similar case, would be when a large number of people open their umbrellas when a rain shower begins (ibid.). This is not a social action for Weber. The individuals are reacting not to other people but to some natural event, i.e., raining. Finally, there is mere imitation. This is when



one is learning to do something by observing a stranger (Greenwood 2003, p. 100). This, for Weber, is not social action because it is only causally determined by the others action and not “meaningfully oriented” to the other persons behaviour (Weber, p. 24).

Social actions, then, must be individual actions meaningfully oriented to others actions. For example, if we are going to walk down the street together, the meaning, or our intentions, must be oriented towards the other doing the same. If we are merely walking down the street next to each other then our actions are not meaningfully oriented towards each other. The logic of this Weberian account is that for two individuals to be acting socially, both need to meaningfully orientate their actions to the other’s meaningful actions.

There are many obscurities in Weber's account and it is unclear if his distinction between social and non-social actions holds on closer inspection. The important point for us is the methodological individualism which underlies Weber's account. He wants to understand how social action is possible without bringing in resources other than the 'meaningful behaviour' of individual actors.

Bratman's planning theory of joint action is a contemporary version of Weber's social action account. The benefit of looking at Bratman's account is its clarity and sophistication while still being wedded to the basic methodological assumptions of Weber's account. By investigating Bratman's account we will see the limits of this individualism as well as see more clearly a fundamental problem with Darwall's account of authority, viz., Darwall's difficulty accounting for how individuals are bound together in not moral cases.

It should be pointed out that Bratman is up front that his account is individualistic. It focuses on:

the shared intentional activities of small, adult groups in the absence of asymmetric authority relations within those groups, and in which the individuals who are participants remain constant over time (Bratman 2014, p. 7).

However, he does think that his account can provide a basis for an extension to

larger institutional agents like corporations or governments, institutions with hierarchical authority relations, with potential flux in the list of their members, and, perhaps with an embedded distinction between those participants who are officials of the institution and those who are not (ibid., p. 8).

We will see two fundamental objections to this modest sociality strategy. First, there is the problem of persistence. How is it that individual persons, as free and equal, are bound together in shared cooperative activity? By looking closely at this issue, it will become evident that individualistic accounts face a dilemma. Either, they will not get beyond small group activities to institutions which necessarily have normative force, or, they will need to bring in resources from outside of their account which will be less parsimonious than a non-individualist account – parsimony being a major reason one would want an individualist account in the first place.

Second, there is the problem of the 'own-action', or mineness, condition that seems to be a necessary condition for intentional states. This is a question about how it is possible for one to practically reason for another, something that seems to be happening in cases of authority in that authoritative commands are meant to bring closure to one's practical reasoning as well as give the addressee a new normative reason for action. Before turning to these two issues, we need to set out more clearly what Bratman's account looks like.

Bratman gives us 3 conditions for joint action in his "Shared Intention Thesis":

We intend to *J* if and only if:

1. (a) I intend that we *J* and (b) you intend that we *J*.
2. I intend that we *J* in accordance and because of 1a, 1b, and meshing subplans of 1a and 1b; you intend that we *J* in accordance with and because of 1a, 1b and meshing subplans of 1a and 1b.
3. 1 and 2 are common knowledge between us. (Bratman 1999b, p. 121)

We can see this as an attempt to formalization of what Weber was originally moving

towards. Notice that what holds two individuals together in a Bratman-type shared intention is that both individuals intend that 'we *J*' – both individuals are meaningfully orientated towards the other in terms of the content of their individual intentions.

I do not want to argue that Bratman-like shared intentions are not a real phenomenon. They might very well be an important part of our everyday lives. It is quite possible that if two individuals driving down a road in opposite directions happen upon a small downed tree which is blocking the road that their actions are coordinated in something like Bratman's shared intention thesis (cf. Tuomela 2007, pp. 115-116). The individuals involved are simply intent on solving a small scale practical problem, move the tree so they can continue to drive. Neither can accomplish this alone and can easily be accomplished by cooperation. The idea would be that such one off encounters do not have time or persistence to generate anything like the normativity which we find in social action which are extended in time. I will not claim that Bratman's account cannot explain such small scale interactions. The target here is on the larger scale interactions which Bratman hopes to be able to extend his account to.

In the following, I will be arguing that there is a strong parallel between Bratman's account of joint action and Darwall's account of authority relations. Both begin with individualistic understandings of non-hierarchical relationships and intend to build out from these to hierarchical accounts (Darwall 2013a, p. 167). What I intend to show in the following is that this becomes highly problematic in trying to understand how individuals become bounded to the others in an institution. Bratman thinks that this type of bond “might issue in downstream interactions that induce [...] obligation-based interdependence” (Bratman 2014, p. 72). However, it is unclear how this will happen and why it is the case that 'obligation-based interdependence' should be understood in the way

Bratman suggests. The target here is to look at the individualistic logic behind this account and show why this type of account is inadequate for the relationship of authority which is embedded in institutions. Let us now turn to the problem of bindingness and persistence of joint intentions.

## 2. The persistence Interdependence of Shared Agency and the Problem of Instability

One of the major obstacles of Bratman's account is understand why each individual will continue to intend that 'we *J*'. This is important because, as we have seen, in order for a joint action to occur all participants to the join action must intend that 'we *J*' and have meshing subplans. However, if one participants stops intending that 'we *J*' then the character of the shared intention changes fundamentally. In order to avoid the instability that could possibly arise, Bratman introduces what he calls 'persistence interdependence'. The idea is that each participant of a joint intention knows whether or not the other participants continue to intend that 'we *J*'. In knowing this, they will each adjust their plans accordingly and in "responsiveness to norms of individual plan-theoretic rationality" (Bratman 2014, p. 65). One of the main features of Bratman uses to explain persistence interdependence is the individual norms entailed by the plan-theoretic rationality. Therefore, let us review these norms in order to see how they might create stability and bindingness.

The lesson that action must conform to certain rational norms has been a philosophical theme since Plato and has recently been taken up by many accounts of practical reasoning, including Bratman himself, Christine Korsgaard and Harry Frankfurt. What these accounts are attempting to understand is the rational norms embedded in self-binding of agents. It is through these norms that we see how agents can be effective and

stable. Without affirming a commitment, or set of commitments, there is no sense to being an agent.

We can bring out the force of this line of thought by looking at how we are to make it through a single day. We must plan to complete different activities and fulfil different desires we may have. However, as we all know, there are not enough hours in the day to do everything we may desire or wish to do. This is why the anarchic or wanton soul will never get anything done, will not act in the full-blooded sense.

When a person succumbs to just any desire that comes to them it may feel as if an alien force is pushing her, as movement is imposed upon them 'from the outside'. I take this to be a common experience. We feel pulled by our desires in multiple directions and being thus overcome by a desire we tend to feel alienated from actions – we do not feel that the action performed was really our action, because we cannot identify with the movements of our body. A classic example is that of a drug addict (Frankfurt 1998). The agent may want to stop using, but they feel overcome by their addiction, by a desire that they no longer identify with. Harry Frankfurt helpfully draws this out when he writes:

It is in virtue of this identification and withdrawal, accomplished through the formation of a second-order volition, that the unwilling addict may meaningfully make the analytically puzzling statement that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it (ibid., p. 18).

The addict is “a helpless bystander to the forces that move him.” (ibid., p. 21)

This is superbly illustrated by Eddie from David Rabe's 1984 play *Hurlyburly*. Eddie is a depraved drug fuelled Hollywood wannabe. Throughout the play he continually wants to know 'how things pertain to him'. However, due to his drug fuelled paranoia and 'semantic insanity' it is revealed that there is not really an 'Eddie' there for things to pertain to – that is exactly his problem and why he cannot do anything. We can see this

clearly in the telling response from his friend Bonnie to one of Eddie's nearly incomprehensible rants

about how he is a 'real person':

EDDIE (sitting there, his legs dangling over the edge of the balcony, he is framed in the square of the railing): Talking about love makes you feel like you're watching TV, Bonnie... (noticing the railing framing him, he realizes he looks like a TV image) that why you're so interested? I'm real, Bonnie. I'm not a goddam TV image in from to you, here. (He starts to pound his legs, having a little fit.) This is real. I'm a real person, Bonnie, you know that, right?

Eddie then abruptly demands a sexual act from Bonnie. She responds:

BONNIE (he is reaching for her; she pushes him away and heads down the stairs): You know, if your manner of speech is in any way a reflection of what goes on in your head, Eddie, its a wonder you can tie your shoes. (Rabe 1995, p. 309)

Eddie, is what Frankfurt calls a “wanton” and Plato called a “democratic soul” (Frankfurt 1998, p. 17 and Plato 1997c, 516b). Eddie does not seem to have what Frankfurt has called “second-order volition” (Frankfurt 1998, p. 16) or what we may call intentions or commitments. Rather, he is nothing but a 'mere heap' of desires with nothing to pull him together (Korsgaard 2009b).

If they are able to get something done it will be a purely contingent matter. The wanton will continually chase after whatever happens to win the battle of motivation at a particular moment. In order to avoid this fate, we must prioritise our commitments in the forms of plans in order to coordinate our lives. One will need a general plan (not an exact plan of every movement) in order to coordinate all of one's ends and also help not to get distracted by other inclinations that inevitably crop up.

The importance of this is clear and can be a way of making sense of why intentions are known without observation (Anscombe 2000, p. 14): If I decide to perform a particular action tonight,  $\phi$ -ing, it is necessary (in the practical sense) that I conform my further

reasoning and planning to the constraint 'of me  $\phi$ -ing tonight' (Soteriou 2013, Ch. 13). It is important to put this in this way, i.e., 'me  $\phi$ -ing tonight', in order to mark the distinction between the commitment to  $\phi$ -ing and predicting that 'I will  $\phi$ '. My plan to  $\phi$  tonight is not something I know through observation but rather something I know from the inside because I am the one making the commitment to do so. Another way to make the same point is that our projected actions, characterized by our commitments to  $\phi$ -ing must always be put in the future perfect tense: I will have  $\phi$ -ed (Schutz 1967, Ch. 2.9-2.10). When we are practically reasoning about what to do we are not concerned with the truth of what will happen in the same way as if we were theoretically reasoning about the future. Rather, we must reason about the future as if it were open.

However, it does not follow from this that one cannot change one's own commitments. What we will discover, however, is that because one can maintain rationality in the face of changing one's commitments and the content of one's intentions, this will not be enough of a foundation to build the stability necessary for shared action. After all, shared action, as Bratman understands it, needs to have multiple agents intending that 'we J'. If any of them can change their minds at will then this relationship will be inherently unstable. With these preliminaries in mind, we can not focus on the rational norms themselves.

There are two rational norms implicit in action. These can be formulated, following Bratman, as a 'consistency constraint' and 'means-end coherence' (Bratman 1999a, p. 31). These two constraints are fairly uncontroversial but it is worth saying a word about each. Before turning to characterizing each constraint individually, it is important that these constraints are at least partly constitutive of action as such. If we continually fail to conform ourselves to these constraints we will not be acting irrationally but fail to be an agent at all. Without these norms we will lose the agent. Also, these are not norms of good or bad action

but rather constitutive norms of any action as such, implicitly endorsed by all agents. That is, there will be no fact of the matter as to where the agent stands in relation to the practical questions at stake; there will be no closure on what to do, hence no action.

The first constraint, the consistency constraint, can be formulated as follows: I cannot be committed to two ends that I believe to be mutually inconsistent. It is important to note a distinction in play here. This follows from the distinction between wishing or desiring some end and making a decision to pursue some end, i.e., forming an intention in the planning sense. In both case, we are not entitled to said end. That is, we do not have any prior right to the ends which we wish for or intend. At most we have the right to our means as well as a right to *pursue* our ends.

What we have control over, what makes us autonomous, is the forming of commitments to pursue our ends, not achieving of our aims and goals – this is the root of Aristotle's and Kant's distinction between wish and choice (Aristotle 1984b, 1111b4-30; Kant 1996c, 6:213). This seems also to be the root of Bratman's distinction between desires and intentions (Bratman 1999a, p. 20). We can wish for or desire all sorts of thing but we can only will what 'depends on us' or is 'up-to-us', what is in our power or control. Making sense of this distinction and its implications for agency, Korsgaard writes,

to will an end is not just to cause it, or even to allow an impulse in me to operate as its cause, but, so to speak, to consciously pick up the reins, and make *myself* the cause of the end (Korsgaard 2008, p. 59).

What we have control over then, what our autonomy is essentially about, is the 'picking up the reins', as it were. There are many things we cause incidentally or accidentally. It does not make sense to demand reasons for the behaviour which caused them. For example, what reason did one have for alerting a prowler who happens to be in the kitchen when one goes for a glass of water in the middle of the night (Davidson 2001, pp. 4-5)? None. The reasons



for turning on the light, rather, was to get a glass of water. The getting a glass of water is what rationalizes the turning on the light. The alerting the prowler was only incidental to this decision. Or to put it another way, what we have control over is making up our minds about what to do and initiating the actions necessary to bring about those ends.

It is only when we make a decision or form an intention to pursue some end that we become rationally committed to the end. Once one has formed an intention is it that the consistence constraint is in play. We can desire or wish for all the mutually inconsistent ends we want – what would stop us? However, once we commit ourselves to some particular end, we cannot, on grounds of irrationality, also commit ourselves to another end which is inconsistent. Why? Because we would no longer be able to fulfil what we have set ourselves to fulfil. We cannot effectively coordinate our future behaviour so that it will be in line with our past decisions. Notice here that we have, in a sense, an 'I-thou' relationship (Korsgaard 2007). The 'I' of  $t_1$  makes a normative claim on the same 'I' at  $t_2$ . Notice that the inconsistency only arises when one believes there to be one.

To briefly illustrate, let us imagine a case in which I commit myself to two mutually inconsistent ends:

C1: I will bring a book to the office today to read.

C2: I will not carry anything on my walk to the office today because I want to have an enjoyable walk and carrying things makes walks less enjoyable.

These are obviously inconsistent. C1 commits me to carry something to the office today and C2 makes it impossible for me to do so. It is irrational to knowingly maintain both of these commitments. I cannot make a coherent plan to do both at the same time.

The second constraint, means-ends coherence, means that I must fill in my plan with further with 'sub-plans', i.e., means, that will bring about the end which I have committed myself to. For example, if I make a decision to  $\phi$  tonight then I must not make

an incompatible decision to  $\psi$  and I must conform any further practical reasoning to the constraint of 'me  $\phi$ -ing tonight' (Soteriou 2013, Ch. 13). That is, I must conform to and reason in accordance with the self-imposed constraint of me  $\phi$ -ing tonight. I cannot know that the means will bring about the ends which I intend and still not intend those means. If I fail to this, I ought to have a self-referential reactive attitude towards myself, e.g., regret. I ought to regret that I did not live up to my commitment to  $\phi$ .

There is nothing said so far that would stop me from rescinding my commitment to  $\phi$ . For if I can bind and commit myself to a certain end, it is not, strictly speaking out of the question that I change that end. What we have in these types of cases is a tension between two 'I's, the past 'I' – in an I-thou relationship – which made the commitment and the present 'I' which must determine whether it will grant authority to the past 'I' and regard the commitment as valid. There seems to be at least four primary reasons why the future 'I' might not take up the commitment of the past 'I'.

First, as has already been mentioned, making incompatible commitments. I cannot, on pain of irrationality, commit myself both to  $\phi$ -ing and to  $\psi$ -ing when these commitments directly contradict one another. I cannot (rationally speaking) be both committed to travelling and committed to stay at home during the same period of time. I may of course be indecisive or desire to do both but once the commitment is formed in favour of  $\phi$ -ing, it becomes irrational to then commit to  $\psi$ -ing. It is not impossible to actually commit to two incompatible commitments. Surely, this has happened to everyone. However, if it comes to my attention that I have two incompatible commitments, I ought to be motivated to drop one of them. If I fail to rescind one of these incompatible commitments then I continue to run afoul of the consistency constraint, hence I am being irrational.

Second, if I change my belief about what is entailed by a particular commitment.

For example, I commitment myself to defending my honour in a duel. Then, I have an epiphany: “If I win the duel I will be murdering someone! Murdering is a bad thing to do!” Based on this epiphany, I realize my end entails a second end which I was already committed to rejecting, viz., murder. I now must rescind one of these commitments. Following John Broome's discussion of wide-scope rationality (Broome 2007), either I have have to give up the intention or have to give up the belief about the action.

Third, one might realize that they have not done enough to really bring about the end which they have committed themselves to. This is to run afoul of the 'means-end coherence' constraint. However, it seems open to an agent to rescind the commitment in order to maintain their rationality. We might realize that the cost of maintaining the commitment has become too high for us to bother continuing to commit ourselves to the end. I might intend to go for a walk tomorrow so I make a commitment to myself to go for a walk in the morning. However, I slept in too late, have a lot of work still to accomplish and it is raining out. It seems reasonable to just rescind the commitment to go for a walk today. If someone continues to do such things they may never accomplish anything. However, there seems nothing said so far which would make one locally irrational about periodically rescinding a commitment.

Finally, there seems to be no principle reason why someone could not change one's mind. Why would there be? Yes, it might be costly for a person to rescind a commitment but surely that is the prerogative of a free agent. If after stepping back from and reflecting on one's commitments, surely the agent can decide not to maintain this particular commitment any longer. This would be the case with an 'I' which made a commitment at  $t_1$  but no longer sees the point in pursuing the commitment at  $t_2$ . Since the past 'I' is in the past it is hard to see how such a change of heart by the 'I'  $t_2$  at could be contested by the past 'I'.

Perhaps there needs to be something that motivates the change of heart but this does not have to be because of either of the constraints on rational action. There also maybe more 'fundamental' commitments that are more resistant to changing but this surely will not be the case for most or even all of our commitments and intentions.

In all these cases we have no reason to regret the rescinding of the commitment (this does not mean we will not have some feeling of regret, rather, it means we have no normative reason to regret) except perhaps the third case but it is by no means obvious that in every case we should. In order for rescinding these commitments in this way to be irrational, we would need to give an independent account of why rescinding any commitment is irrational.

Now the point of going through these reason one can rescind a commitment is that even when a commitment is rescinded by an individual agent it does not render that agent ineffective. The agent carries on, fulfilling other commitments. As we will see, this is not so in the case of shared agents. The instability is a much greater problem for shared agents then it is from individual agents. With this in mind, let us turn to case of two individuals in an I-Thou relationship and the creation of modest sociality.

The idea of the persistence interdependence is to account for why the intentions of each participant in a shared action continues intending that 'we  $\phi$ ' and how they mutually reinforce each other in so intending. To see how this works it will help to have an example in front of us. I will use a recent example from Margaret Gilbert and discussed by Bratman (Gilbert 2014, Ch. 5; Bratman 2014, pp.116-117). To put the example in Bratman's terms:

1. (a) Ned intends that we hike to the top of the hill (b) Oliver intends that we hike to the top of the hill.
2. Ned intends that we hike to the top of the hill in accordance and because of 1a, 1b, and meshing subplans of 1a and 1b; Oliver intends that we hike to the top of the hill

in accordance with and because of 1a, 1b and meshing subplans of 1a and 1b.  
3. 1 and 2 are common knowledge between them.

Now obviously there might be all sorts of important subplans to be settled by Ned and Oliver. We can assume that this all gets worked out between them. However, what will happen if Ned changes his mind? According to Bratman's own account, the shared intention no longer exists:

[O]nce Ned has changed his mind they no longer have a *shared intention* to climb to the top, one that is set to explain their climbing to the top. After all, at that point Ned no longer intends to climb to the top, or that they together climb to the top, and instead fully intends *not* to climb to the top; and what is now going to need to be explained is not their climbing to the top, but rather their *failure* to climb to the top (Bratman 2014 p. 117).

This makes any shared intention to be contingent on each individual's continuing to share the content of the intention. So long as all the participants of the shared intention maintain their commitment to that content, in this case hiking up the hill, the shared intention exists. However, once one individual stops sharing the content of the intention, the shared intention is fundamentally changed and might cease to exist at all. This is because "shared intention need not ensure mutual obligations" (Bratman 1999b, p. 134).

Bratman, however, thinks that even though shared intentions are contingent on each individual maintaining their individual intentions that 'we  $\phi$ ', that there is reason to think that this is strong enough. This is because each person's intentions, Ned's and Oliver's, give each other "mutual rational support" (Bratman 2014, p. 70).

This rational support comes in three potential forms: desirability-based, feasibility-based and obligation-based interdependence. However, this last form of persistence interdependence is drawn from outside of the structure of shared intentions, for example, in the form of making promises (ibid., p. 72). We will return to this last form in a moment. For now, we will focus on the first two forms.

The first two types of persistence interdependence are quite straight forward. For desirability-based interdependence, Ned and Oliver both judge that it is desirable to continue to intend to hike up the hill as long as the other continues to so intend but if either stops intending to hike up the hill the other would no longer find it desirable to so intend (ibid., p. 70). They only want to hike up the hill if the other also wants to hike up the hill.

Feasibility-based interdependence is structurally similar (ibid, pp. 71-72). However, opposed to only finding it desirable to continuing to intend if the other continues to intend, Oliver sees Ned's intention to hike up the hill as what makes his hiking up the hill feasible. Perhaps Ned is the only available person to drive Oliver to the hill. If Ned no longer intends to hike up the hill, it might no longer be possible for Oliver to get to the hill. Likewise, Oliver is the only of the two who knows where the hill is, if Oliver changes his mind then Ned will no longer find it feasible to hike up the hill because he will not know how to get to it.

If this account is going to be able to account for larger, more stable institutions then these two types of persistent conditions seem to be inadequate. For example, the types of institutions that characterise the modern state or universities or workplaces etc which are the type of institutions which have authority relationships embedded within them, seem to need something like obligation-based persistence conditions.

Bratman at this points turns to Scanlon's 'Principle F' which says "If A provides B assurance that she will do x, in the absence of a special justification, A must do x unless B consents to x's not being done" (Hindriks 2013, p. 476; cf. Bratman 1999b, Ch. 7 and Scanlon 2003). However, we can ask, 'who are we obligated to'? If the obligation is only to ourselves, it looks like the other party really has no reason to hold us accountable for changing our minds.

What is wrong with Scanlon's 'Principle F' and why doesn't it create stability which we are looking for in larger groups and institutions? First, we can ask why expectations have any normative force on its own. We typically do not attach normative force to expectations to the behaviour of things that are not human. Furthermore, there are many cases in which we have a positive evaluation of someone's actions precisely when they do not conform to our expectations. Expectations as such do not seem to have any normative force. Now Scanlon would be quick to point out that 'Principle F' only gives *pro tanto* reasons to conform to expectations. So it might be that in the cases just mentioned that we have over-riding reasons to look past exceptions.

Another reason to have doubts about the principle would be that the normative force that we find in institutions is not about what we are obligated to do individually. The obligations are directed obligations rather than obligations of how one ought to act. In this way, they are not individual obligations. As Gilbert puts it:

Let us assume that a given promisor will have an obligation – indeed, a moral obligation – that derives from Scanlon's Principle F. It is not at all obvious that this obligation corresponds to a right of the promisee against the promisor to performance of the promise. In order for it to do so, it will have to be not just an obligation, but an obligation towards the promisee, an obligation that is the other side of the coin from the promisee's right against the promisor to performance. An obligation [...] is not, or not necessarily, a directed obligation. (Gilbert 2014, pp. 278-279)

If our target is to understand authority, this type of obligation will not do. As we have seen, the question of standing needs to be answered with a relational, directed obligation. The authority needs to be able to hold the addressee accountable. Thus far, Bratman's persistence conditions, even his obligation-based condition, does not seem to bring about the type of bond necessary to make sense of authority or the institutions which they are embedded.

This is the instability that is at the heart of acting together when conceived solely in

terms of an 'I-thou' account. We might want to invoke something like Scanlon's 'Principle F' in order to overcome this instability. However, it does not seem that this principle can do the required work. In the case of individual commitments both individuals have individual intentions with the content of some end in which the other is a necessary component, to be sure. Further, they might have deeply held (and accurate) expectations of what the others commitments are and how they will act. However, in the case of acting together, there is a need for both parties to be united in their commitment in a different way than is necessary for cases of expectations.

### 3. Normative Construction of Agency, Practical Identity and the Bindingness of Commitments

So far we have looked at Bratman's theory of planning agency and seen that nothing in synchronic norms of rationality impose a strong form of bindingness on agents. That is, as far as Bratman's theory is concerned, no irrationality in changing one's individual intention to  $\phi$ . However, this is not to say that there is not a bindingness of individual intentions and commitments. Bratman does have a further strategy which he develops in order to secure the stability of intentions over time.

What Bratman is trying to deal with in his account of the diachronic aspect of agency is how to answer two questions. First, the problem of subjective normative authority. This is about the relation between desiring or having a pro-attitude towards some end and taking oneself to have a normative reason to pursue that end (Bratman 2007, pp. 90-91). Reflective distance is one way this space can open up. When we step-back from our desires we are able to deliberate about the normative reasons for and against pursuing some desired end. The deliberation takes the form of 'should I endorse this desire?'. Second, the problem of agential authority. This is about how to make sense of the difference between



cases of motivated action, i.e., a desire has lead to an action, and cases of action which were governed or determined by the agent (ibid. pp. 92-93).

Bratman's answer to these two questions is through his planning theory of action. That is, agents commit themselves to prior, future directed plans. These plans, according to Bratman, have a certain form of stability in that "there is, normally, rational pressure not to reconsider and/or abandon a prior plan" (ibid., p. 26). Initially in *Intention, Plans, and Practical Reason* Bratman develops some ideas about the diachronic stability of agency over time (Bratman 1999a, Ch. 5-6). This is what Bratman in his later work calls the 'intention stability strategy' (Bratman 2007, p. 264). These ideas revolved largely around the reconsideration of plans. Stability, Bratman argued, was achieved because reconsideration is both costly and risks undermining coordination. Furthermore, the deliberative resources exerted when one reconsiders a plan can reduce the effectiveness of temporally extended and limited agents.

Of course, Bratman does not want to say that reconsideration of plans is generally irrational (Bratman 2012, p. 79; cf. Bratman 2007, p. 26). There are many cases in which it is perfectly rational to reconsider one's plans. For example, one might acquire new information that undermines one's earlier plans or one might recognize that the formation of one's earlier plan was flawed in some way (ibid.).

These general consideration leads Bratman to introduce the following diachronic rationality constraint:

The following is locally irrational: Intending at  $t_1$  to  $X$  at  $t_2$ ; throughout  $t_1$ - $t_2$  confidently taking one's relevant grounds adequately to support this very intention; and yet at  $t_2$  newly abandoning this intention to  $X$  at  $t_2$  (ibid.).

The idea being that unless there is some reason to reconsider one's intention at a later time, one is rationally constrained not to reconsider one's prior intention. That is, if one does not

acquire new information or recognize some flaw that grounds the prior intention, one should remain committed to one's plans.

However, as Bratman recognises, this is a fairly minimal constraint (ibid., p. 80). The constraint is relaxed enough that one can change one's mind simply because it “would be a nice change of pace” (ibid.). This hardly seems like much of a constraint at all. If something as minimal as it would be nice for a change is enough to maintain rationality when changing one's mind then it is unclear when, if ever, one would run afoul of such a constraint. Furthermore, much of the time when we change our minds there are reasons to do so. For example, if I am tired, is that enough reason to reconsider an earlier plan? If so, there does not seem to be much binding one to one's prior intention.

Bratman has a further response to attempt to deal with his bindingness problem. This is what he calls the 'agential authority strategy' (Bratman 2007, p. 265). This strategy revolves around anchoring the stability of intentions in the agent's point of view. The idea being that the agent can gain reflective distance from a current desire and decide not to endorse such a desire. Since this endorsed desire is endorsed by the agent of having a certain value, according to Bratman, it becomes a general policy that is embedded in the point of view of the agent (ibid., pp. 271-275). The idea is that this general policy has priority over singular intentions and desires of the agent that crop up because it is the general policy which structures the ongoing practical reasoning of the agent. The general policy structures the ongoing practical reasoning of an agent because it is an intention about “the weighting of pros and cons in one's motivationally effective practical reasoning” (ibid., p. 273).

The idea here is that if I have a general policy to walk to the office everyday rather than take the bus then as I practically reason about how to get to the office today I will give

more weight to walking than taking the bus. This is true even if today I feel like walking. The general policy, according to Bratman, gives more weight to the plan of walking than taking the bus which should outweigh my current desire to take the bus. However, again, why should I not just decide to change my general policy? It is unclear what gives the general policy can not just be changed as particular intentions can be as well.

The final point Bratman makes about stability of intentions is to look at the stability within intentions themselves. He argues that when we form an intention to  $\phi$  that we must conform to the 'no-regret condition' (Bratman 1999b, p. 79). The 'no-regret condition' says roughly that if we were to stick with our intention to  $\phi$  in the future we will be glad and if we do not we will regret that we did not. As he notes, "this no-regret condition includes both the absence of regret at having followed through and the presence of regret if one did not follow through" (ibid., p. 79n37). However, if one has a reason to change one's mind, why should one feel regret? They may regret changing their mind but that seems to be a contingent issue about a particular person's physiological make-up. It is not obvious why one should regret changing one's mind.

The issue seems to be with Bratman's psychologism about commitments. Following Jonathan Dancy, "[p]sychologism is a view about motivation; it is the claim that the reasons for which we act are psychological states of ourselves" (Dancy 2000, p. 99). One point in which Bratman makes his psychologism explicit is in the following remark: "we need to know whether this phenomenon of agent (or, self-) determination consists in some, perhaps complex, causal structure involving events, states, and processes of a sort we might appeal to within a broadly naturalistic psychology" (Bratman 2007, pp. 91-92; cf. pp. 28-32, 99-100, 262). If, as I have argued, the stability of commitments cannot be convincingly achieved through Bratman's account of agency, we will need to look for another model.

The competing model of agency we will look at is the one the Christine Korsgaard has developed. This “normative constitutionalist” model of agency (Korsgaard 2014) has a more promising answer to the stability question, or so I will argue. This strategy focuses primarily on the question of agential authority but gives a different and more compelling answer than that of Bratman's in that it does not rely on a psychologistic understanding of commitments but rather a normative understanding.

Korsgaard aligns this model with four of the most important names in the history of philosophy: Plato, Aristotle, Rousseau, and Kant, as well as with the contemporary work of Margaret Gilbert. I think we can add others that might fit broadly into the Kantian tradition, such as Hegel's practical philosophy (Hegel 2008; cf. Pippin 2008, especially Ch. 3) and Robert Brandom's understanding of the norms of practical inference (Brandom 1998 Ch. 4, Brandom 2000 Ch. 2, Brandom 2009 Ch. 2-3, Brandom 2013).

According to the normative constitutionalist view of agency, the agents’ ability to create laws or norms for herself transforms mere activity into the robust action of an autonomous agent. In a rough characterization of this view of agents, Korsgaard tells us:

These philosophers suppose that the kind of unity that is essential to the notion of agency is something that must be normatively constituted, that is, something that is achieved by conformity to certain norms. (Korsgaard 2014, p. 192)

Importantly, the norms which must be conformed to for the unity to be achieved are the norms created by the agent.

The theory contains the idea that there is a normative relationship between the 'parts' of the agent, i.e., the multiplicity of desires that tend to pull agents in multiple directions. One part of the agent speaks for the whole in the sense that it imposes/creates the norms which unify the parts. This is how Korsgaard understands the agential authority of imposing norms. However, we can not step back from our desires and choose to endorse

them from a view from nowhere, Korsgaard claims. Rather, we must reflect on our desires and impose norms from a conception of oneself, from a practical identity “under which you find your life to be worth living and your actions to be worth undertaking” (Korsgaard 1996a p. 101). However, for Korsgaard, agents typically have several practical identities which, she claims, are all grounded in the fundamental identity of as human. We can leave aside this grounding project of Korsgaard's. What will be important for us here is how practical identities bring stability and unity to an agent in a stronger form than was available in Bratman's planning theory.

Notice, however, the similarity to Bratman's account, particularly his claims about general policies and the weighing up of pros and cons in ongoing deliberation. There is a similar move here on Korsgaard's part. However, there are some fundamental difference. Namely, for Korsgaard a practical identity is a much deeper phenomenon than a general policy. As Bratman describes it, a general policy seems to be just about particular types of activities, how many glasses of wine, for example. Korsgaard's practical identities are about being a certain type of person and the integrity inherent in one's identity. To understand why this is, we need to look more closely at Korsgaard's account.

Korsgaard has helpfully pointed to Plato's tripartite distinction in the soul so as to highlight this relation: Reason is the head of a man, Appetite a multi-headed monster, and High-Spiritedness comes in the shape of a lion head (Plato 1997c, 588c-e; Korsgaard 2009 Ch. 6-7 and Korsgaard 2008, Ch. 3.2). Alternatively, we might think of the alternate metaphor Plato uses in the *Phaedrus*, in which the charioteer (Reason) holds the reins of two horses, one of a noble breed (Spirit), and one of an ignoble breed (Appetite).

Plato's metaphor of the multi-headed beast seems particularly apt for the idea of appetite. Appetite can be fruitfully understood as equivalent to Kant's *inclinations*, which

are “[h]abitual sensible desire[s]” (Kant 2007, 7:251) or “[t]he dependence of the faculty of desire upon feelings is called inclination, and this accordingly always indicates a *need*” (Kant 1996b, 4:413note). Desire and appetite pull us in multiple directions. Most of us have, at least at times, felt as if we are being torn apart by different and conflicting desires we want to fulfil: ‘Should I be a lawyer or a philosopher?’ What is needed for agency on this view, then, is a way to bring the disunity or diversity driving the potential agents in multiple directions together into a unity. This is necessary in order to legitimately attribute an action to an agent. If any inclination wins through its will-to-power, the inclination, rather than the agent, is in control. Furthermore, it is the norms which hold the agent together that are the reasons that the agent has and are therefore the reasons that the agent is accountable for. This is why people need the unifying power of a rational will, to which we will now turn.

Our Reason – our will – allows us to have a practical identity due to Reason’s nature as the legislator that gives direction to the appetite through rational commitments. Reason is identified as the part of the soul that has the standing to speak for the whole. Whether we account for these in terms of some sort of ‘second-order volition’ or as ‘reasons’, we will see that this is how we can make sense, not only of the *person* acting but also how this connects with the idea of accountability. We can see this as connected to *willing* in the Kantian sense:

duty is the necessity of an action from respect for the law [...] Now, an action from duty is to separate off entirely the influence of inclination, and with it every object of the will; thus nothing remains for the will that could determine it except, objectively, the *law* and, subjectively, *pure respect* for this practical law, and hence the maxim of complying with such a law, even if it infringes on all my inclinations (Kant 1996b, 4:400 – 4:401).

How should we think about the legislative power of the will? The will must be thought

about in procedural terms so as to avoid the problem of disunity. If we are to give substantive ends to our will – that is, if the will must be subordinated to the inclinations – the will is not autonomous. The will would be dictated by something outside of itself and the agent would no longer be able to say that it has given itself its own laws or norms (Kant 1996b, 4:441). This does not mean that inclinations are not important for agents. Rather, they serve as in-puts into our practical reasoning, in-puts that we must endorse (or not) through the process of reasoning itself – they ought not determine what a good outcome of the reasoning must be. This is achieved through reflective distance.

Consider the way Korsgaard describes the way reflective distance effects the relationship between persons and the norms which grip them. She says that “Self consciousness opens up a space between the incentive and the response, a space of what I call reflective distance” (Korsgaard 2009, p. 116). What is at issue for us is no longer just how to achieve our ends, to do this first than that, how to gobble up the objects of our desire, but our very ends themselves – should this be done at all. This reflective distances loosens the grip which desires have on us. John McDowell makes a similar point. He says,

One difference reason would make is to bring the facts about what wolves need to conceptual awareness, and so make them available to serve as rational considerations. But what converts what animals of one's species need into potential rational considerations is precisely what enables a rational animal to step back and view those considerations from a critical standpoint. So when they become potential reasons, their status as reasons is, by the same token, open to question [...] Reason does not just open our eyes to our nature, as members of the animal species we belong to; it also enables and even obliges us to step back from it, in a way that puts its bearing on our practical problems into question (McDowell 1998, p. 172 and §10)

In other words, due to our ability to step back from our desires, we put those very desires into question. At this point we are unable to plan and carry out our actions. Action for a person, is always a matter of accepting a norm of action for oneself. Making sense of our actions and valuings. That is, we posit a norm for our own action and we are responsible to

conform to those norms by recognising them. In this way, our actions are intelligible to ourselves and to others. Valuing and desires, for Korsgaard, only become reasons when we recognize them and endorse them as our reasons for action.

Let us look at an example of this process which will help in understanding what is going on. We can use the modern hero who embodies such and is trapped by reflection, Hamlet. His famous soliloquy in the first scene of Act Three ends with these lines:

Thus conscience does make cowards of us all;  
And thus the native hue of resolution  
Is sicklied o'er with the pale cast of thought,  
And enterprises of great pith and moment  
With this regard their currents turn awry,  
And lose the name of action. (Shakespeare 1992, Act III Scene i)

In this soliloquy Hamlet is deliberating the question: "Whether 't is nobler" to endure the rot that is Claudius "Or to take arms against a sea of troubles" (ibid.). However, thinking too much has led him to contemplate failure. I take this to be a paradigmatic problem in deliberating about what to do.

Hamlet's, as well as our own, thinking can imagine more possibilities, more potential points of view that he could endorse than he could possibly carry out the actions which that would follow from endorsing them – should he take the standpoint of prudence or of vengeance? When deliberating we are not only concerned with different conflicting and incompatible ends to realise but also with the person we are to be and our integrity to our identities. Each option puts different commitments into question; each option puts us into a position in which we must abandon some other end we may value. Yet, the more we think the more we cannot act. "We may say" to quote Theodore Adorno, "in general that this discrepancy, this divergence of consciousness and action constitutes the central theme of *Hamlet*" (Adorno 2001, p. 112; cf. Adorno 2006, p. 231f and Adorno 1973, p. 228). This



is exactly what we might expect if we take seriously the idea of reflective distance. We step back and ask, are these rational ends to pursue, should I really value this? What accounts for this inability to act, the losing the name of action, once we gain the ability to step back from our desires to deliberate about which ends should count as reasons for us? For Hamlet it is his deciding which hero to be, for us it is deciding on which practical identity to endorse.

Following Korsgaard, I think the answer is a disunity of the soul which is the question of stability. We begin to split apart contemplating these different options, these different selves we could become. That is, how we are going to constitute ourselves. Obviously, Hamlet's problem is more extreme than most. He seems to be going mad. The point seems to be a question of whether this action is the type of action Hamlets want to characterise himself. The literary critic Harold Bloom captures this when, as he writes about the end of the play, that Hamlet “does not die as a vicarious atonement for us, but rather with the single anxiety of bearing a wounded name. Whether we ourselves expect annihilation or resurrection, we are likely to end caring about our name” (Bloom 2001, p. 217).

We can be struck by key moments in our lives where we are uncertain not only of what it is we are to do but who we are to become. When we encounter certain obstacle and problematic situations we find ourselves forced to step-back and ask who we are. We do not simply have doubts about our means to realise what we will but the ends in which we will as well. Doubt arises as to whether it makes sense to be this type of person or not. The point is that the question of who we are, or the question of self-constitution, is deeper than Bratman's content question about what kind of policy we should have about wine drinking. Rather, what Korsgaard is pointing to is the identity, the character, of the one creating the

policy. Internal to this identity, stability is usually secured by integrity which is what is challenged in these difficult moments of deliberation – to lose one's identity is to lose oneself.

We are looking for a way to maintain the integrity of our will in difficult situations which we confront. Constitutive of this integrity is asking what type of will we ought to have. The rightness is about maintaining ourselves in the face of disunity. If we are trying to decide between becoming a lawyer or philosopher we are trying to decide what type of person to be, to figure out who we are, since we cannot be both or at least most of us cannot be both. As Korsgaard puts it “to will an end is not just to cause it, or even to allow an impulse in me to operate as its cause, but, so to speak, to consciously pick up the reins, and make *myself* the cause of the end” (Korsgaard 2008, p. 59). The anxiety about rightness then is not outside of us; it is about making ourselves actual through our actions. This is the circle of self-constitution.

Hamlet begins in uncertainty. The conclusion of his deliberation is a commitment of his will; it is him determining what his ends *ought* to be. What rule of action he is going to lay down for himself. The conclusion of practical deliberation then is a commitment of the will. It is deciding what to do, deciding upon a course of action.

This is about putting constraints on our own action. When we treat a consideration as a reason, we represent the world in a certain way by recognising those self-imposed constraints. In this way, to treat those self-imposed constraints on our action as true is to act in a way that respects those self-imposed constraints. When we impose a constraint ourselves.

Conceptually tied up with the legislative activity of the will is an accountability to conform to the norms given by Reason. Besides being necessary to maintain the unity of

the agent, this second aspect is important as it is the success criterion of agency. Only when an agent conforms to its self-willed norms is the agent successful. As Brandom has recently noted:

The distinction that action implies, between purpose and achievement is in play because these are the elements one must compare in order to assess success or failure. And the unity essential to the concept of action – the fact that endorsing a purpose, adopting it as one's own is committing oneself to a norm according to which the achievement *ought* to be what one intends – is just what sets the normative standard for success. Disparity of purpose and achievement is failure (in accomplishing what one intended to accomplish); identity of purpose and achievement is success (in accomplishing what one intended to accomplish) (Brandom 2013, p. 78).

The idea is that integrity of practical identities is more than just extra motivation of plans. It brings with it a normative criterion of success. If we fail to live up to the self-imposed norms which we endorse from a particular identity, or standpoint, we ought to regret our failure. We ought to have this reactive attitude towards ourselves. In this way we are accountable to ourselves for our failure to do as we willed, to do as one of our practical identities gave us reason to act. This does not, of course, make it impossible to change our minds, to change who we want to be. However, it does make us normatively responsible for doing so and we ought to have the reactive attitude of regret for failing to live up to our own norms.

#### 4. Own-Action Condition and the Mineness of Intentions

Let us now turn to the second problem with Bratman's shared intention account. Its violation of the own action condition. That is, it violates the condition that says that “it is always true that the *subject* of an intention is the *intended agent* of the intended action” (Bratman 2014, p. 13). Bratman himself acknowledges that his account violates this condition but he goes on to reject such a condition (ibid.). Before turning to Bratman's rejection of the own-action condition, let us look at why it is an intuitive condition to

accept.

A good place to start is with Michael Oakeshott and something he writes in his last and under appreciated 1975 work, *On Human Conduct*. He writes there that the “definitive postulate of human conduct” is “free agency” or “self-determined autonomous human beings” (Oakeshott 1975, pp. 234-235, 315 and Ch. I.II.2; cf. Friedman 1989). In other words, free agents are autonomous in the sense that they have the capacity to decide on which commitments to make and which projects to pursue. This type of autonomy does not beg the question of authority in the direction of philosophical anarchism, contra R.P. Wolff (Wolff 1970). We are not here dealing with political or moral authority but rather with personal autonomy; we are dealing with a central aspect of the metaphysics of human agency.

Another way to put the idea is: only a particular type of minded being can be an agent in a robust sense. “Let us say that a person is *minded* in a certain way, if he has the perceptions of salience, routes of interest, feelings of naturalness in following a rule, etc. that constitute being part of a certain form of life” (Lear 1982, p. 385 and Lear 1984, p. 229). A being can only be an agent in a robust sense if the agent can engage in practical reasoning. Which only minded beings are able to do. Rather, minimally, for something to be an agent there must be some reasoning about the means to its ends; something cannot be determined otherwise what sense can be made of agency?

Now it should be clear why practical reason is important here – at least in a minimum sense. The ability to make simple choices in a plan is a necessary condition to attribute an action to something and hence for something to be an agent.

Consequently, the intention to do this (rather than that) must be that of the one who is acting in order for it to be the one agent who brings about, to be the cause of, the end.

Others seem to only be able to encourage or influence our actions. They cannot, however, make up our minds, intend, or act for us. As Hans Bernhard Schmid writes, evidently with Heidegger in mind:

[j]ust as one cannot die the death of others, even though in some cases, one can die *for* them, one cannot pursue the other's goals without making these goals one's own. This is an essential fact about our intentionality (Schmid 2009, p. 124).

This dimension of intentional agency is the mineness (*Jemeinigkeit*) of intentionality (Heidegger 1962, p. 42). For your action to be your action and not being moved by some external force you must be the one picking up the reins. One of the most important elements of this mineness is that it is what makes us accountable for the actions in which we perform. We are accountable for our actions *because* they are our actions. They are the actions which we decided to under take and make ourselves the efficient cause of.

This fact has been pointed out by many in objection to Michael Bratman's theory of joint action (Velleman 2000, p. 203; Schmid 2009, p. 124). The objection runs as follows: if we conceptualized joint intentions as 'I intend that we  $\phi$ ', then each person who is part of the joint action must intend the other's  $\phi$ -ing. However, how can it be that one person can settle a practical question for another? It is not in one's power to intend the  $\phi$ -ing of another.

Bratman responds that we do not need to accept the own-action condition because it is quite normal for us to 'intend that' others  $\phi$  which is different from 'intending to' (Bratman 2014, p. 60). The idea seems to be that indeed the own-action condition does attach to 'intending to' but we often 'intend that' others do things. For example, parents intend many things for their children: *that* they go to a good school, *that* they do their homework, *that* they go to bed on time etc. Or a teacher might intend that her student are prepared for the seminar. However, this seems to miss the point of the own-action condition. When we intend that our children or students do something, it does not seem that

we are committing them to anything. Rather, we are only trying to influence their behaviour. We want them to take into account our desires for them in their own practical reasoning. Our 'intentions that' they do something does not settle what they will be doing. Only they can do that.

Another important aspect of the own-action condition, as Frederick Stoutland has noted, is that “[t]he own-action condition is implicit in another constitutive condition: an agent cannot intend to A if she is not prepared to take full responsibility for having done A intentionally – even if doing A depended on the will of others (or on luck)” (Stoutland 2007, p. 241). The way that Bratman has interpreted this is that separating out the 'two faces' of intentionality, that between psychological explanation and understanding and taking responsibility – to focus on the latter is to moralize the verb 'to intend' (Bratman 2014, pp. 62-63). Let us note two points about this move.

First, by making this move, Bratman shows that he is not entitled to any of the bindingness that comes from practical identities and commitments of the will that we have looked at in the previous section. This is important in that one might think that an appeal to practical identities might lend enough bindingness to shared intentions to account for the stability of institutions (Laden 2000).

Second, it is not entirely clear that being 'prepared to take full responsibility' is necessarily a moralizing move. One possible way to understand Stoutland (although it is unclear if he would accept this) would be to think of the responsibility in a de-moralized way. What one is responsible for is the rationality of one's actions. In other words, one must be prepared to take full responsibility for Bratman's own rationality conditions. When we intend for others, it does not seem that we are prepared to take full responsibility for the rationality of their actions. Rather, we are only concerned with a limited part of their

psychical economy, viz., what we intend for them to do. When we ourselves are acting, we must be concerned and responsible for the whole psychical economy otherwise we would be liable to run afoul of the consistency constraint on intentional action.

To turn back to the problem of authority, it does seem that this is what authorities are trying to do when they give commands. They are intending and practically reasoning for their addressees. We can see this in two respects. First, a command brings to an end the deliberation of the addressees by giving them a new first-order reason for action and a pre-emptive reason for not considering other reasons (see Chapter One §§4-5). Second, the addressees are accountable for following the intentions of the authority (when the authority is legitimate). If they do not conform to the intentions of the authority, they are doing something wrong. They are accountable to the authority for their conformity.

Therefore, it seems that we will need an account of authority which can account for how authorities can practically reason for their addressees without conflicting with the own-action condition.

## 5. Robust Sociality: Grounding Authority in Social Practice

To conclude this chapter and Part I, let us summarise. We now have two criteria needed to establish the standing of authority. First, we need to be able to account for the bindingness between the authority and the addressee. Our critique of Darwall's conception of authority has opened up the way to see how to approach accountability and from there how to begin to answer the question of standing. What we have learned by looking at the inadequacy of Darwall's account is that starting with individual commitments of autonomous individuals leads us directly into the problems of the particularity requirement. It is unclear how we are to determine who is accountable to whom if the autonomous

individual has the right to rescind her commitments when she chooses. Further, if we cash-out the commitment that is apparently incurred when a commitment is formed, we find that the obligation is not directed which is part of Darwall's point but it is unclear how he can make such a point with the way his theory is structured. The obligation is not to the other person but rather an obligation about which acts are acceptable.

The second criterion is the own-action condition. That is, we need an account which can overcome the tension between the autonomy of individual agents and the necessity of intentions being the intentions of the subject who is acting intentionally. It is not clear at all that Darwall has any resources in his account to deal with this problem either.

In order to avoid this fate, we should turn our investigation to institutions. This will be the topic of Part II. There I will argue by understanding the social ontology of institutions we can fulfil both of these criteria. Furthermore, by understanding institutions in this way, it also allows us to answer both the question of standing (Chapter Five) and the question of content (Chapter Six). Before turning to these questions, we will first need to say more about social ontology and how it will meet the two criteria for standing discussed in this chapter (Chapter Four).



## **Part II: Governing Together: On the Social Ontology of Authority**

## **Four: *On the Bindingness of Social Ontology***

Part I has left us with questions about how binding institutions are. In order to answer this question, it will be helpful to step back and ask a prior one which John Searle has posed: “What is the ontology, the mode of existence, of social institutions?” (Searle 2005, p. 1). In this chapter, I am going to argue that the way we need to answer this question is to move away from the individualism and overcoming the 'Cartesian brainwashing' of contemporary theories (Schmid 2009 Ch. 2 and 9; Pettit). The way to do this is to develop an adequate theory of institutions and collectives. It is in these social relationships that authority emerges. What is typically missing from contemporary theories of legitimacy, and what I want to suggest is the reason for the inadequacies found in them, is the recognition of the authorities as embedded in larger institutions. Many classical political theorists from Hobbes to Rousseau and Hegel\* have had some sort of holistic understanding of the state. As Christian List and Philip Pettit suggest, “as fascism took over Europe, however, [holistic theories] also became associated with a totalitarian image of society, and this may have led to its ultimate demise” (List and Pettit 2011, p.9; cf. Gilbert 1992, p. 428). Without a full understanding of the social ontology of institutions it is difficult to see why institutions should not be thought of as simply apparatus of coordination between monadic individuals. This is how aggregative democrats understand the functioning of democracy, for example. It is hard to see how or why anyone must necessarily volunteer to join and maintain a state apparatus which may not always serve her narrow, or even wide, self-interest. The problem with individualistic accounts are obvious: the authority of the state cannot necessarily be reduced to the preferences and benefits of

\* Not to mention Marx 2000, Simmel 1910, Durkheim 2013, Dewey 1983 and 2008, Ch 3 and Mead 1962

the individual. However, without a proper understanding of institutions is difficult to see an alternative.

The idea here has echoes of Rousseau's remark that: "[t]he commitments which bind us to the social body are obligatory only because they are mutual, and their nature is such that in fulfilling them one cannot work for others without also working for oneself" (Rousseau 1997, p. 61). What we are going to need to determine in the following is how to understand the relationship between individual agents and institutions. That is, we need to provide an account of the social ontology which makes it intelligible as to why individual agents are bound together in institutions. Understanding how individuals are bound together will be the key to determine an answer to the question of standing and the question of content.

## 1. The Social Ontology of Institutions

A helpful place to start is to follow Searle a bit further by making a distinction between something being ontologically objective and something being ontologically subjective. The rough distinction goes as follows. What is ontologically objective is anything that could exist independently of human minds: atoms, mountains, the sun etc. Ontologically subjective, on the other hand, is anything that could only exist in virtue of the activities of humans: art, buildings, money. This distinction is the basis of all social ontology. Everything that needs to be accounted for in terms of social ontology will be ontologically subjective. However, it is not necessarily the case that everything that is ontologically subjective also requires a social ontology. We might think, for example, that at least some things are ontologically subjective but not social phenomena. They might only require one mind for their existence. So they would be ontologically subjective but not

social. Perhaps some individual tastes or desires are not social in this way but still are ontologically subjective (however, I do not want to commit to a view on this particular issue here).

It should be obvious that what we are concerned with here is the ontologically subjective and the social. If there were no human beings there would be no institutions or authorities and, presumably, if there were no human beings needing and wanting to coordinate socially there would be no institutions or authority either. By definition, then, the way to account for the mode of existence of institutions is to give a social ontological account.

However, this distinction, although fundamental, is still not enough to get us to a proper social ontology. This calls for a further distinction, one between holism and individualism. It is important to be clear what is and what is not meant by holism. There is one view of holism which is beyond the scope of the present discussion. That is the social holism of Charles Taylor and Philip Pettit, among others (Taylor 1985, Ch. 7, Pettit 1993, Ch 4, Pettit 2002). Their concern is with certain human capacities, for example, the ability to think, have language or be self-conscious, which they claim can only be achieved through interaction with others. Now clearly there would not be the human capacity to think if there were no humans. However, these social holists make the stronger claim that in order for one individual to, say, think it is necessary for that individual to interact with other individuals. They are making a claim about the logical impossibility of a congenital Robinson Crusoe to think or have language.

The holism of the present discussion consists in a more modest claim. It is the claim that to understand collective social phenomena we cannot appeal only to the actions, beliefs, preferences etc. of individuals (Gilbert 2000, p. 155). Accounts that only appeal

individual actions, beliefs, and preferences are individualist accounts of social ontology. Bratman and Darwall are individualist in this sense. This chapter will serve as an introduction and a justification for a holistic account of social ontology. In chapters Five and Six this account will be used as the basis for answers to our two questions for authority.

It is important to note the ontological distinction between individualism and holism does not have any necessary consequences for a similar epistemic distinction (Searle 1995, pp. 5-13; Searle 1998 pp. 44-45; Searle 2005, §2; Searle 2010, pp. 17-18). That is, something can be ontologically subjective and yet it does not necessarily follow that it is epistemically subjective. Institutional facts, while being ontologically subjective, are still epistemically objective. The facts about institutions are, in some sense, independent of the beliefs, desires and actions of individuals within an institution. This is something that needs to be explained by an account of social ontology.

Yet, this gives rise to a puzzle. If institutions could only exist in virtue of activities of human beings, then how can institutions still be epistemically objective in that the facts about these institutions are independent of the individual beliefs, desires and actions of those who make up the institutions? Are we not facing the spectre of a group-mind? That is, are we not in danger of ascribing intentional states to a super-individual entity? The worry here is that we typically think that “[o]nly individuals have minds and only minds can have intentional states” (Hindriks 2003, p. 217; cf. Schmid 2009, Ch. II.7). Many may be troubled by such an ontological implication of holistic accounts. As we shall see, this is a misguided worry as the leading accounts in the contemporary literature are all committed to denying the existence of group-minds in any mysterious sense. This is because there is also room for an 'acceptable ontological holism' (Gilbert 1992, p. 431), one in which there is no super-individual entity over and above individuals but only individuals who construct

social collectives.

If we were to follow Bratman's thesis about shared intentions and where to build an account of institutions from it then it seems that we could not make sense of this puzzle. For Bratman, the institutional facts are indeed dependent on the individual intentions. As we have seen, if participants in a shared action change their minds then the nature of the shared intentions, for Bratman, by necessity must also change. The facts of an institution on Bratman's account then must be dependent, perhaps in some complex way, on facts about the particular mental states of members making up the institution.

Yet, institutions are not like this and Bratman's ambition to extend his account to larger groups “with potential flux in the list of their members” (Bratman 2014, p. 8) betrays this fact. Institutions as we normally understand them tend to have a stability about them which is independent of the mental states of the 'list of their members'. It is this that lets us say that even though institutions are ontologically subjective they are still epistemologically objective. How do we make sense of this?

## 2. The Problem of Collectivity: The Necessity for a Social Ontology of Institutions

It is becoming generally recognised that there are forms of human collectivity that cannot be understood by adding traits individuals together. There are clear cases in which human collectives can be accounted for by mere addition. For example, there are human collection of right and left handed people. These collections all depend on individual attributes of the members that make them up. This 'adding' of traits is what has been called a summative account (Quinton 1975, p. 9, 17;cf. Gilbert 1992, p. 19). We can see the simple phenomenon in our daily lives. However there are several standard examples that cannot be accounted for through a summative account: painting a house together, planning

a trip together, dancing together and walking together. In other words, in order to understand any institution we need to understand what kind of collective these are and why a summative account is inadequate.

The reason why a non-summative account is necessary to explain these phenomena is that they present peculiar features which actions of individual agents do not present. They involve many agents non-accidentally performing an action together. If two individuals accidentally perform an action at the same time they are merely acting in parallel. There is a 'togetherness' or collectivity that would be missing and needs to be accounted for making the performance non-accidental.

There have been two types of argument given for the need to move towards a holistic social ontology in order to understand the collectiveness of cooperative activities. The first type of argument has been developed in response to issues in game theory and the necessity to move to what has been called 'team reasoning' in order to account for the cooperation actually observed (Sugden 1993). However, since there is no commitment in the present work to a rational choice model of individual agency we can leave this argument aside.

The second type of argument is stronger and does not presuppose a rational choice theoretic framework, viz., the conceptual argument given by Raimo Tuomela. He writes:

There is a conceptual and logical difference between saying that a) I have as my goal that we together paint the house [...] and that b) I have as my goal to participate in our painting the house. However, for the “thick” notion of participation by jointly intentionally acting together it holds on conceptual grounds that a) is true if and only if b) is. The thick notion of participation entails the participants' collective commitment to the collective action in question (Tuomela 2000, p. 32).

The idea which Tuomela is getting at is that if our target phenomenon is of doing something together, then we have to conceptually understand what it means to participate in a

collective or to do something together.

Take, for example, two musicians playing music. There are (at least) three descriptions of a scene where two musicians play in harmony. First, the two musicians could, perhaps implausibly, be unaware of each other and happen, because they picked the same piece of music play in harmony. There is no collectivity here.

Second, there is Weberian 'social action' or Bratman-type shared intention where the two musicians know that the other musician is going to be playing music at the same time and they each intends to play in harmony. Perhaps they are able to coordinate their behaviour by following a certain convention. The convention might be: "always play along with whatever the first player is already playing". In this case, each orientates her behaviour towards what the other is doing; the musicians then have reached a 'coordination equilibrium': neither "one would have been better off had *any one* agent acted otherwise" (Lewis 2002, p. 14). Even though this case is non-accidental action it seems more appropriate to call it acting in parallel rather than acting together. There still seems to be something missing, a 'togetherness'.

It will take us a bit of work to get to the third description which will account for this togetherness. One might suspect that a summative account maybe able to account for this 'togetherness'. However, if we are to account for togetherness we must be aware of a particular problem in a summative account. In many cases, the content of individual intentions diverge. It is unclear how a summative account, which adds up the intentions of individuals, can account for such divergence. As Searle as made the point:

There may be a difference between the content of an intention or a belief of a group as a whole, and that of an individual member of that group. The claim is that the difference in content reveals that there is a genuine difference between collective intentional states and individual intentional states (Searle 1990, p. 403).



To carry on with the musicians playing together, it maybe that the group intention (however that is accounted for) is to play Debussy's *Petite Suite*. However, because this piece is written for four hands on the piano, each individual will have to play her own part; each individual's intentions will be about playing this particular part. Neither individual will have as the content of her individual intention to play Debussy's *Petite Suite*.

However, this is still a case where the intentions combine in a neat way. We might think that we can account for this type of collectivity in some complex summative account. There are cases in which this cannot be done.

The cases I have in mind are ones that meet what Margaret Gilbert calls the 'disjunction criterion' (Gilbert 2014, p. 102). That is, when two individuals form a collective, neither of the individuals need have an individual intention whose content is part of the overall collective intention. "There can be a tension or even outright conflict between what one intends or believes personally and the intention or belief one has as a group member" (Gilbert 1996, pp. 201-202). The idea is that one or even both of our piano players might change their individual intentions. Neither may want to play Debussy. Perhaps they both, individually want to play Dvořák's *Legends* but neither has communicated this to the other. However, their actions will still be coordinated and they can play Debussy together even though neither personally wants to play the Debussy piece. What keeps them acting together is not their individual intentions but the further phenomenon which we are trying to characterise.

Evidence for this being the case is to think about what would happen if one of our musicians started to play the Dvořák piece rather than the Debussy piece. The other musician would rightly rebuke her. She might say something like "we were going to play Debussy, why are you play Dvořák?" Or simply, "that is not what we were going to play!"

If nothing else, there would certainly be a moment of surprise on the part of the second musician as it was not the intention of the two musicians as a group to play Dvořák.

Or it might be that both of them play the Debussy part together and afterwards one turns to the other and say, 'I really did not want to play that piece. Next time we should play Dvořák's *Legends*.' The other musician at this point might reply, "I feel the same way! I had no intention to play Debussy but that was what we were planning on playing." Again, the point of these examples, is that the individual intentions of the musicians need not match what the two of them, as a group, intend to do and yet they can both successfully play the piece they intended to play.

Also, any plausible account of collectives and in particular institutions need to allow for the possibility of radical misunderstanding (Gilbert 1992, Ch. IV §4.1(6)). For example, a group of people may agree to write a letter of protest together. However, even though they have agreed to write this letter together, they may not have realised that the content of their objections are fundamentally different. Yet, for the time being, they still seem to be a collective whose content is to write a protest letter together. While they have a common understanding of what they want to do and accomplish, they do not have a consensus. As Charles Taylor has remarked:

common meanings are quite other than consensus, for they can subsist with a high degree of cleavage; this is what happens when common meaning comes to be lived and understood differently by different groups in a society. It remains a common meaning, because there is the reference point which is the common purpose, aspiration, celebration. Such is for example the American Way, or freedom as understood in the USA. But this common meaning is differently articulated by different groups. This is the basis of the bitterest fights in a society, and this we are also seeing in the USA today (Taylor 1985, p. 39; cf. Gilbert 1992, p. 213).

The way to make sense of this phenomenon, I will suggest, has to do with the ability of individual members of a collective to rescind or change the commitments of the group

unilaterally.

In other words, it is possible that because there is such a cleavage in common meaning between our letter writers that some of them no longer wish to participate. Yet, they are committed until they as a collective rescind the commitment to write the letter. This is a further criterion of collectives and particularly important for institutions. It is precisely this bindingness that is inherent in the formation of collective enterprises like institutions.

### 3. The Bindingness of Institutions

In order to see the connection between collectivity and bindingness, let us take up a historical example from Kant's *Metaphysics of Morals*. Kant worries about how it is that one can make a contract with someone and at the same time maintain both individual's autonomy. If both individuals are autonomous and have the right to their own means and property, neither can take from the other. However, a transfer cannot rightfully take place if one party gives up what properly belongs to them. This would be abandoning, not a proper transfer. Rather, a transfer can only properly take place according to Kant:

“through a common will by means of which the object is always under the control of one or the other, since as one gives up his share in this common undertaking the object becomes the other's through his acceptance of it” (Kant 1996c, 6:271).

This unification, Kant conjectures, is symbolised in the shaking of hands (ibid. 6:272).

Kant gives a helpful example to clarify this point, viz. the act of promising. He tells us that “what belongs to the promisor does not pass to the promisee (as acceptant) by the *separate* will of either but only by the *united will* of both, and consequently only insofar as both wills are declared *simultaneously*” (ibid.). That is, when we make promises it is not that one person makes a personal commitment and then waits for the other person to accept the commitment. There is a gap here that must be filled. Rather, the obligations which come about through a promise apply to both parties simultaneously. This will initially strike some as counter-intuitive because we sometimes think that if I promise you something, you can unilaterally relinquish the promise, that I can only be let out of the promise if you consent. This is not, in fact, the case; at least, not in such a simplistic form. We can see this if we look closely at how the institution of promising actually works. By doing so, we can also see that Kant's point seems to bear out.

If  $A$  promises to meet  $B$  to  $\phi$ , neither  $A$  nor  $B$  have an obligation until both have an obligation. Let us take the promise to be meeting you tomorrow. If  $A$  says to  $B$ , “I promise that I will meet you tomorrow at 2pm”, it is clear that  $A$  is not obligated until  $B$  accepts. This is clear but is also what misleads many into thinking that promises can occur between an 'I' and a 'thou' without the forming of a united will. However, think of what actually occurs. Why is it that  $B$ 's acknowledgement of  $A$ 's promise puts  $A$  under an obligation? Does  $A$  commit herself and then feel relieved that  $B$  accepts? No. The commitment only occurs when both  $A$  and  $B$  come together as a single 'we'. This has further implications.  $B$  also incurs obligations: the obligation to be at the appropriate place, at the appropriate time. It seems clear that if  $A$  shows up at 2pm and  $B$  is nowhere to be found,  $A$  will be justified in rebuking  $B$ . This seems to be not only because  $B$  has an important and indispensable role in enabling  $A$  to fulfil the promise but because  $B$  is obligated to  $A$  as well.

Kant further elaborates what he has in mind in the *Metaphysics of Morals*, both with the example of sexual relationship and marriage which he characterises as “a relation of *equality* of possession, equality both in their possession of each other as persons [...] and also equality in their possession of material goods” (ibid. 6:278). Here, we see Kant struggle to overcome the basic dilemma of authority. How is it that we are able to subject our will to another if our will is autonomous? Kant's answer seems to be that, in order for this to occur, there must be a union of wills, the creation of a 'we' where both 'I's retrieve their autonomy by mutually, and reciprocally, giving themselves to the other. This works because if I possess you and everything that comes with you, and vice versa, I also possess myself because that is something which you possess. Kant's remarks on this issue are more or less schematic but I think is generally on the right track.

Later, Hegel took up Kant's main idea and was able to expand and deepen it.

Although Hegel thought “Kant presents marriage in a shameful, ugly manner” (quoted in Williams 1998, p. 215), this seems to be a verbal dispute. Take, for example, the following passage for the *Phenomenology of Spirit*:

[T]his movement of self-consciousness in relation to another self-consciousness has in this way been represented as the action of *one* self-consciousness, but this action of the one has itself the double significance of being both its own action and the action of the other as well. [...] Thus the movement is simply the double movement of the two self-consciousnesses. Each sees the *other* do the same as it does; each does itself what it demands of the other, and therefore also does what it does only in so far as the other does the same. Action by one side only would be useless because what is to happen can only be brought about by both. Thus the action has a double significance not only because it is directed against itself as well as against the other, but also because it is indivisibly the action of one as well as the other (Hegel 1977, §182-183; cf. Honneth 2012 Ch. 1).

Here, Hegel is making the point that the desire for recognition from another can only be had through the unification of both individual self-consciousnesses into a 'we' – Hegel later elaborates this point in the third part of *Philosophy of Right* as *Sittlichkeit* or 'ethical life' where membership of a community is of fundamental importance. Both acting together jointly, parties maintain themselves as unique, autonomous individuals by both allowing the other to be through self-imposing constraints on their own freedom. This idea is too complex to go into detail here (see Williams 1998, Pippin 2008, Brandom 2007). However, the point is the same, though in a more general way, as Kant's observation on promises and marriage that the possibility for their occurrence necessitates a double moment of the 'I' in the 'we' and the 'we' in the 'I'. These two historical examples were not meant to convince. Rather, they serve as models of how to overcome the dilemma of authority by understanding how individuals can be bound together as a united will and yet still maintain their autonomy.

In the case of the 'we', if we are committed to some end then each individual, qua group member, is responsible for doing their part in actualising the end and is accountable

to the group for non-compliance, i.e. not doing their part. The group's prior intention, then, constrains the individual's future actions and reasoning. The question of authority then becomes about how roles can be formed which give individual members the power to determine the group's commitments. In this way, authority is derived from the group itself – it is properly our authority or the authority of us – and the ends to which it is committed, not particular individuals over other individuals. This seems to be a natural way to think about positions of authority, particularly when we find something problematic about the individual who happens to be in a particular position of authority, when we, in Tocqueville's phrase, respect "the office rather than the official" (Tocqueville 2003, p. 237).

If this is right then the problem of authority as set out above dissolves. That is, if we can conceive of commands as commitments of a group which have been delegated to a certain role, then the infringement on autonomy is no longer a problem and the legitimation of the coercive force of commands is also answered. The former problem is answered because the autonomy of the individual is preserved by being committed to the institution in which authority is embedded. The commitment is still, in a sense, the commitment of particular individuals, as we will see. It is not an individual or personal commitment of the individual but another type of commitment which the individual makes, a joint commitment. The answer to the latter problem would run something like this: just as in the individual case, one must remain true to one's commitments, i.e. to follow through with the decisions one has made, so in the group case we must remain true to our commitments. Part of the mechanism to achieve this is necessarily going to consist of having critical reactions towards non-compliance in order to bring one's inclinations in-line with the desired goals and ends. That is, if I fail in trying to reach my ends then I should be regretful or angry with myself. In this way, I will hopefully have more of an incentive to achieve my aims next

time around or there will be inertia provided by my 'high spiritedness' or disposition. The same holds true for the institutional case. If I do not fulfil my role within the institution then I am open to rebuke, not only from myself qua member but also from other members.

The kind of bindingness we are looking to understand is going to have to be more than some associative accounts of obligation claim (which was first discussed in Chapter One, Section 1). As was observed in the first chapter, the idea behind this type of account is to establish similarity between the special relationship which holds between parents and children and the special relationship between citizens and their state. The idea is that both of these relationships give rise to obligations and duties; that membership in and of itself entails obligations. Both Dworkin's and Horton's accounts, in different ways, attempt to provide conditions to establish what membership means.

For example, Dworkin in *Law's Empire* writes about the connection between community and obligation in the following terms:

the members of a group must by and large hold certain attitudes about the responsibilities they owe one another if these responsibilities are to count as genuine fraternal obligations (Dworkin 1986, p. 199).

For Dworkin, each individual needs to hold certain attitudes. It is from these attitudes that each individual actually understands that an obligation has arisen. This becomes more apparent when Dworkin lists four conditions on associative obligations: “regarding the group's obligations as *special*”; “accept personal responsibility”; “see these responsibilities flowing from a more general responsibility each has of *concern* for the well-being of others in the group”; “*equal* concern for all members.” (ibid. pp. 199-201).

However, it is not just that one holds these attitudes but rather that if one is a member of such a community, what Dworkin calls a 'true community', one must hold these attitudes. In other words, if one is formally a member of a true community then one does



have the obligations. It is type of community which only gives rise to genuine obligations.

For Horton, membership must be ‘acknowledged’. This acknowledgement is ‘subjective’ in the sense that it relates to a more or less *explicit* self-understanding, incorporating moral sentiments, emotions and attitudes” (Horton 2007, p. 12 *emphasis added*). However, Horton makes a similar move to that of Dworkin when he claims that associative obligations can only arise from “associations that can have value for its members” (Horton 2010, p. 176).

What both of these accounts hold in common is that the strength or the bindingness of the association is not actually located in the association itself. The association part of their accounts is an attempt to answer the particularity problem. However, in each case the association itself is not what brings the bindingness of institutions, or communities, out. First, there seem to be communities and then, if the community is judged to have value or judged to be a true community, then and only then is there bindingness (cf. Knowles 2009, p. 190).

If this is right, it seems unclear from instrumental accounts what the substantive difference is between the source of obligation and the source of bindingness. That is, both accounts seem to rely more on the importance of equal concern or having value for the particular members. It is not clear what the binding work of associative obligations is when rendered in this way. The binding work seems to be done more on these content-full concerns about the associations rather than the associations themselves (cf. Simmons 2001, 79n35).

For this reason, we should attempt to move away from this way of formulating the idea of associate obligation and attempt to understand the bindingness of association and the bindingness of institutions for an internal point of view. We should attempt to

understand, if possible, how being part of such an institution can on its own bind us to the institution. As we will see in Chapter Six, this does not have the consequence that every command within that institution is valid, which might be part of the problem that Dworkin and Horton are trying to avoid, but first we need to make more sense of the bindingness of institutions.

#### 4. The Constitution of Social Practices: We-Mode Accounts and Subject Accounts

So far, we have been discussing social ontology and institutions at a rather high level and showing why individualistic accounts are problematic. What we have found is that we need to account for several different things in a proper understanding of institutions. First, we need to understand how it is possible that individuals' attitudes can differ from the facts about the social institution which they in some way help to constitute. This is important because if we fail to account for this we either end up with a super-individual group mind or some complex, but still inadequate, summative account.

We also need an account that can make sense of the bindingness that seems to be inherent in the formation of institutions. This is for two reasons. First, many people feel a sense of belonging and obligation to the institutions to which they belong. However, it should be clear from this that we cannot look for a psychological account which is "a mere matter of 'feelings'" (Hart 2012, p. 56) of belonging to a certain collective. Rather, what these attitudes point to is a (potential) aspect of something deeper, viz. a social rule. What we want to know is whether these feelings are an error or if they rest on a deeper social rule. As H.L.A. Hart makes clear, "Such feelings are neither necessary nor sufficient for the existence of 'binding' rules. There is no contradiction in saying that people accept certain rules but experience no feelings of compulsion" (ibid.). What we are looking for is: what

gives an institution their binding power which accounts for the standing of particular authorities? We want to know whether this is an error on members' parts or if their feelings rest on deeper social rules of institutions. Second, as we saw with the examples of Kant and Hegel, there is an important sense in which we are bound to each other in institutions which give us roles to perform which, if we fail to do, we violate those to whom we are bound.

Currently in the literature there are two broad alternatives to Bratman's individualistic account of shared intentions. First, there are Searle's and Tuomela's 'we-mode' accounts (Searle 1995 and 2010, Tuomela 2010 and 2013). Although Searle and Tuomela differ on how they account for the 'we-mode', they both share the idea that proper understanding of collective social phenomena is through the 'mode' of intentions. Where their accounts differ is in how to account for 'we-mode' intentions. Searle claims that the we-mode or we-intentions are primitive natural phenomena (Searle 1995; Searle 2010). Tuomela attempts to give a reductionist account of them to more basic intentions (Tuomela 2007; Tuomela 2013). The we-mode account postulates that there are two different modes that individual subjects can have as an intention: the I-mode intention and the we-mode intention. Notice the 'we-intention' is attributed properly to the individual mind: I we-intend to  $\phi$ .

It seems, however, that the we-mode account is going to run into the same problems as Bratman's content-based approach. There does not seem to be any clear way in which there is any binding to be had in an individual with a we-mode intention. The reason why we should be sceptical about such a we-mode account is that it is not clear why there is bindingness on the individual to the group in these cases. What would stop someone from either changing their we-mode intention to an I-mode intention? Or from simply rescinding the intention all together?

A more promising account is the one purposed by Margaret Gilbert. Gilbert's theory is a commitment account of collective phenomena. As she puts it:

I refer to populations as “collectives” when I conceive of them as genuinely collective subjects of intention, action and so on. I take it that a population is a *genuinely collective* subject of intention if and only if, roughly, it can plausibly be regarded as having an intention *of its own*, an intention, if you like, of the population *as a whole* (Gilbert 2014, p. 236)

In this joint-commitment account of collectives, it is the 'we' of a group who commit as a whole to believe or intend something. It is the 'us' as a collective who is the subject of the intentional state. These plural subjects, or 'we's', are formed through what Gilbert calls 'joint commitments'. I think this is a promising account for understanding both the bindingness of institutions as well as the way to overcome the own-action condition. We shall look at this account and its applicability to the issue of authority in the final two chapters.

With these basics on social ontology now in place, we can turn to our main target: to understand how authority is embedded in institutions and how this helps to provide answers to our two questions for authority. It is time now to turn back to these questions for authority.

## ***Five: Social Ontology and the Question of Standing: The Owing Account***

The first two chapters of Part I left us with two questions which need to be answered to determine the legitimacy of authorities: the question of content and the question of standing. There is good reason to think that the question of standing should be seen as the proper place to start when attempting to give a full account of legitimate authority. This is because it is a necessary condition of authoritative commands to be possible at all. Without standing to give commands the question of content, the question of 'are these commands valid?' seems to be irrelevant. The command itself does not even 'make it' to the addressee in the proper way. Without the proper background relationship in place, the addressee is not accountable to the authority. This then gives the question of standing priority. We must determine how the relationship of authority to their addressees is possible. If the argument from Chapter Three is right, then we cannot adequately address the question of standing within an individual methodology. We need to turn to the holistic account that was outlined in Chapter Four. That is, the plural subject account offered by Margaret Gilbert.

The claim is that the two general problems with establishing standing can be overcome through understanding the social ontology of institutions. Remember, the problems were how it is possible to acknowledge that human agents are necessarily autonomous in the sense of both the own-action condition and how addressees are bound to particular authorities. The basis of both of these problems stem from a conception of individuals as self-originating sources of valid claims. This is an intuitive idea of how to understand the importance of individual persons. It is the basis of both of these problems in the following way. First, it is the source of the own-action condition because for an agent to

act at all they must be the ones intending, it must be their self-originating claims that underlie what they are doing. More important, however, is the bindingness condition. The problem with accounting for bindingness is that, if an individual is a self-originating source of valid claims, then why should they not be able to change their minds at will? What is to stop them from leaving an institution when they decide that they no longer want to be a part of it? That is, what is to stop them from rescinding commitment with no normative reason for regret? If we cannot overcome these two general problems, then we have to face the same issue as Darwall's account, viz. the problem of the would-be independent.

In the following I argue that, to overcome these two general problems, it is important to look at authority at the right level. That is, it seems to me that these general problems only arise if one assumes that the only conceivable level of understanding agents and authority is at the individual level. The problem of authority is often posed as a matter of how *A* can command *B*: how can one individual command another isolated individual? This is precisely why it is troubling to think of commands in terms of the intention in the name of another. It is unclear how or whether this is even possible. Another way we can put this, which highlights the implausibility, is to ask: 'how can one practically reason for another?'

This individualistic model of the social ontology of practical authority has the tendency to conceptualise authority relationships only between separate beings. However, this tendency is often hidden because the question is put in terms of how states (as corporate entities) can command individual citizens, for example. Yet the logic of the answer has it that the question is still thought of in terms of individual *A* commanding individual *B*.

This, I submit, is the predominant way of posing the problem of authority

descending from the consent tradition. Under this conception, authority, when legitimate, is typically seen as being transferred from one individual to another. You have authority over me if, and only if, I consent to your authority. This is an attempt to overcome the problem of 'mineness' of intentions and the own-action condition. Yet, as we have seen, this ends in instability because of the inability to account for the bindingness of such transference. If one can make a commitment to obey some authority, what stops them from having a change of heart?

It does not seem to me, however, to be the problem of authority with which many other traditions have been concerned. Among those who want to ask the question differently are Hobbes, Rousseau, Kant and Hegel. They seem more interested in the question of self-governance in the first person plural: how do *we* govern ourselves? Although this account has conceptual obstacles of its own, notice that the problems of 'mineness' of intentions and bindingness are not among them. Rather, what authorities do is settle our deliberative questions for us: the subject of the intention is the same as the intended agent of the intentional action. In other words, *A* inhabits the role in this institution to settle these deliberative questions for the institution as a whole in which *B* is a member.

It is important to notice how distinct a question this really is. In the individualistic version, we face the problem of transferring authority to overcome the problem of practically reasoning for another: How does one give up one's natural authority to another? When does this transfer actually take place? Why are the individuals not entitled to change their mind?

In this alternative, holistic model, we do not have these problems. This way of conceptualising the relationship of authority is one in which the authority is created or, better, emerges between individuals when they unite in institutions. The concern for this

model is: how are institutions formed which can give rise to the roles of authorities and addressees?

The reason for the move to the holistic model can be put in a straightforward way: if we accept the idea that as individual agents we can form commitments to pursue certain ends then, if we form institutions, we can also legitimately form commitments to pursue certain ends as a group – we can 'pick up the reins' together. How does this solve our problem? In order to be an autonomous agent, when I commit myself to an end then I am responsible for actualising (or trying to actualise) that end as well as being accountable at least to myself if I fail. In this way, my prior intentions constrain my future actions and practical reasoning. In the holistic model, this autonomous agent is the plural subject of an institution: we commit ourselves to an end and are responsible for actualisation of the said end.

## 1. Standing and Authority as Owing Relation

It is now time to turn to understanding the strong form of binding that is involved in institutions. This will allow us to understand how authorities can have standing to give commands and hold their addressees accountable for conformity.

In the process, we will see that there is some truth in the consent tradition. There does indeed seem to be a story about what gives someone a right to rule and why individuals are committed to their particular institutions; we need to explain why individuals are bound to their institutions. However, opposed to the story of the transfer of authority via consent, authority is constituted by the creation of institutions by us and the obligation to defer is owed to the institution as an us.

Although we want to understand the authority in larger institutions like those found



in states, it is helpful to try to understand authority on a dyadic level first for the sake of simplicity. Obviously, we then need an argument as to why this dyadic mode of authority can be translated to macro-level, but this can wait until we have a better grasp on what is to be translated.

How should we understand the relationship between individuals and their institutions? The best place to start is with a proposal Margaret Gilbert has made. She maintains that the appropriate standing to command can be at “least partially explicated” by reference to the “owing relation” (Gilbert 2006b, p. 247). In her words,

if X has the standing to command Y to do A, then, when X issues to Y the imperative 'Do A!' or does something that amounts to this, Y owes it to X to comply with this imperative (ibid. p. 248; cf. Gilbert 2014, pp. 414-415)

To use a simple example to illustrate: Bill owes Nathan a favour. If Nathan issues to Bill the imperative “Pick me up from work today!” Bill owes it to Nathan to pick him up from work. Furthermore, if Bill refuses, then Nathan is in a position to rebuke him for non-compliance. Gilbert has suggested that this is not only a necessary condition for Nathan to have standing over Bill but is also “at least close to a sufficient condition” (ibid.). This seems plausible as a first approximation.

Notice, however, that we have started with some form of background owing relationship. In the simple case above, it was the background relationship of Bill owing Nathan a favour. It is this background relationship which we need to explicate as it is the foundation of authority. I argue that this background relationship is best understood as a member of an institution. Being a member of an institution, in turn, is best understood in terms of joint commitments.

## 2. The Social Ontology of Authority: Joint Commitment as Foundation

How then does this background relationship of standing characterised as an owing relationship come about? I follow Gilbert in her answer: through the establishment of a joint commitment. To understand this answer, we need to know what a joint commitment is and how it is formed. Gilbert understands the term as:

A joint commitment... is a commitment of two or more parties. It is not a combination of commitments, one of one party, one of another, and so on. Given their joint commitment, each party has sufficient reason to act accordingly, just as one has sufficient reason to act according to a personal decision one has made. As [Gilbert] understand[s] the phrase, if one has sufficient reason to do something, then one is rationally required to do it, all else being equal (Gilbert 2006a, p. 158, cf. Gilbert 2006b Ch. 7 and Gilbert 1996 Ch. 6) .

The first thing to see is how this diverges from the accounts of sociality we have so far considered. First, unlike Bratman's individualistic account, this is clearly a holistic account. Where Bratman's shared intention thesis focused on the individuals' intentions and how they combine, Gilbert is interested in understanding the jointness, as it were, of the commitment. Bratman argues that, in order to understand sociality, we should look at how two (or more) individuals' intentions 'that we  $\phi$ ' come together through meshing sub-plans. The structure of Gilbert's account is fundamentally different. She begins by trying to understand how we can commit to  $\phi$ . There is no way, for Gilbert, to understand social phenomena without referring to others.

However, there is also a major difference between the holistic accounts of Searle or Tuomela's we-mode accounts. These we-mode accounts want to understand sociality via each individual's 'we-intentions': I 'we-intend' to  $\phi$  and you 'we-intend' to  $\phi$ . Rather, Gilbert attempts to account for sociality in the subject of the intention, not the mode of the intention. It is fundamentally we, jointly committing together, to  $\phi$  that is the proper foundation of sociality.

Furthermore, the bindingness which we found lacking in the other accounts of sociality which we have looked at is built directly into Gilbert's joint-commitment account. Just as when individuals wish for or commit to a particular end, when we jointly commit to an end there is sufficient reason to pursue that end. Gilbert defines sufficient reason in a standard sense that:

X has *sufficient reason* for performing A if and only if a consideration C that speaks in favour of X's doing A is such that, all else being equal, rationality requires that X do A, given C (Gilbert 2006b, p. 29).

It should be noted that reason here is used in a normative sense, not a motivating sense. That is,

X will act *irrationally* in not doing A if X believes he has sufficient reason to do A, and that all else is equal, and yet does not do A (ibid. p. 29n7).

That is, X has sufficient reason to  $\phi$  in situation S if X is part of a joint commitment to  $\phi$  in situation S, all things being equal. Why do these joint commitments bind us together and obligate each to fulfil their respective roles?

In her essay entitled "Obligation and Joint Commitment", Gilbert lays out five arguments for why a joint commitment carries with it obligations (Gilbert 2000, p. 54-58). Although these arguments are distinct, they generally all come down to the same reason: the "jointness of commitment" (ibid. 55; cf. Gilbert 2006b, p. 163). It is the "jointness" of a commitment that is of the utmost importance for Gilbert. This means that a joint commitment is not the commitment of any of the individuals but rather the commitment of the group itself – it is a 'plural subject' committed to act together as a single body.

The normativity we find here is the same kind of normativity implication of willing we saw in Chapter Three, Section 3. In that section, we looked at how when individuals step back from their desires and commit to certain ends, they are constituting their identity

which brings a form of integrity with it. When we have a particular practical identity, we have sufficient reason to pursue the ends which are entailed by that identity. The same is true here. When a plural subject is created, it gives itself an identity to which all members are bound and have sufficient reasons to pursue the ends which are entailed by that identity. However, instead of this being an individual reason for action, it is a social reason for action. It is a reason for all individuals as members of the institution to act.

We can see now why a joint commitment is strongly connected to obligation and, further, to authority:

By virtue of being party to a joint commitment I *owe my conformity* to the other parties in their capacity as parties. In this capacity, therefore, they all have a *special standing* in relation to my conformity; they have a *right against me* to it, and they will rightly take themselves to have *the standing to demand it* from me *and to rebuke* me if it is not forthcoming (Gilbert 2006b, p. 161).

In other words, the members of a joint commitment have authority to make claims by virtue of joining together.

To understand fully what this means, it is instructive to look closely at how a joint commitment is formed. For this, we need an expression of readiness, i.e. an intention to form a joint commitment as well as common knowledge of this readiness (Gilbert 2006b, p. 138-139). In other words, the expression of readiness needs to be '*out in the open*' (ibid. p. 53; Gilbert 1992, pp. 191-197). The intention to join, for Gilbert, can be "at some possibly quite low level of awareness" (ibid. p. 234) which can be characterised as a 'pooling of the wills.' (Gilbert 1992 Ch. IV §3.7.iii; cf. Gilbert 1996 p. 186).\*

We might worry at this point that an appeal to individual readiness to form a joint commitment will be too individualist or too implausible for our needs. However, Gilbert distinguishes between two types of will formation which help to alleviate this worry. The

\* After *On Social Facts*, Gilbert uses the phrase 'joint commitment' rather than 'pooling of wills', not because 'pooling of wills' is inappropriate but because she finds it "more helpful" (Gilbert 1996 p. 9)

first type is what she calls an “intuitive” exercise of the will, characterised by making a decision. For example, I decide since it is a warm day that I would like to go for a walk with you. So I turn to you and say, “It is really warm out today. Will you go for a walk with me?” You think for a moment and reply, “It is really warm out. A walk would be nice. Yes, I will go for a walk with you.” This is obviously a stilted conversation and it is evident that most ‘pooling of the wills’ does not take this form. Even if some do, it is surely not how most of us find ourselves wrapped up in many of the institutions of which we are a part, particularly the state.

Gilbert's second type of will formation is more helpful in this context. As she explains,

[w]hile a personal decision may be characterized as an *act* of will, a personal *intention* may be characterized rather as a *state* of will, or, to use a common philosophical phrase, a conative state (Gilbert 2006b, p. 128)

In other words, intentions *form* without the active decision of the individual to form the intention that we saw highlighted in the previous example. When an intention to  $\phi$  is shared by multiple individuals to whom this intention is common knowledge, a joint commitment with obligation is formed, or as she sometimes puts it, in this case “a tacit understanding” has “emerged” (ibid. p. 367). Unlike the example used above, the ‘pooling of wills’ is usually this more subtle matter.

The committing of oneself is analogous to what Wittgenstein meant when he said, “they agree in the language they use. That is not agreement in opinions but in form of life” (Wittgenstein 2010, §241). These joint commitments form the background agreements which make intelligible the activities and behaviours of the institution.

One might worry that low level conative states, which are involved in the creation of joint commitments, are not powerful enough to have the socialising effects necessary for

the maintenance and functioning of background agreements (Gilbert 2014, Ch. 11). However, the primary abilities to form such intention develop as early as infancy and the socialising effect can be seen even in this pre-lingual age. For example, Michael Tomasello reports that children as young as 14-months old can understand the background agreements of group activity (Tomasello 2010, §4.1.4). The case Tomasello describes is as follows: an infant (between 14- and 18-months old) and an adult are picking up toys and putting them into a basket together. When the adult points to a target toy, the infant will pick up the toy and put it in the basket. We can gloss the communicative act here as follows: “Pick up *that* toy and put it in the basket.” However, when a second adult, who is not involved in the group activity of picking up toys enters the room and points at a target toy, the infants do not react in the same way, i.e. they did not put the toy in the basket. Tomasello presumes this is “because the second adult had not shared the cleaning-up game as common ground with them” (ibid. p. 127). In other words, the second adult did not have the request relationship which is necessary for joint commitments – the infant did not feel committed with the second adult to pick up toys. It was not intelligible to the infant that the second adult wanted the same thing as the first.

This example demonstrates two things. First, the construction of a background through constitutive joint commitments can be accomplished even by infants – they can understand who is part of a group commitment and how the group commitment constitutes the meaning of future behaviours. Second, and correlated to the first, it shows how constitutive commitments can create meaningful behaviour and different roles which each individual of the commitment is meant to pursue.

We can now make more sense of the Wittgenstein passage quoted above. Constitutive commitments determine a group's identity – form of life – by making certain

behaviours meaningful for the members of the group. Not only are they meaningful but they demand certain other meaningful behaviours in response. It is in this way that these commitments constitute a non-reducible 'we', a form of life.

It is no accident that institutions and language can be explicated in such similar ways. As Peter Winch once remarked:

[t]here are important formal analogies between language and other institutions; for to act in context of an institution is always to commit oneself in some way for the future: a notion for which the notion of being committed by what one says provides an important parallel” (Winch 1972, p. 70; cf Searle 1995, p. 59, Searle 2010 Ch. 4 and Tomasello Ch. 5).

Hegel made a similar point when he said that “we can see language as the existence of Spirit. Language is self-consciousness existing *for others...*” (Hegel 1977, §652; Houlgate 2013, p. 169). That is, in language we become publicly committed with others in that we must recognise ourselves and others as mutually committed to a shared public meaning.

One example of this more subtle form of intention being formed is through repetition. You and I can become jointly committed in the following way: after the departmental seminar, we decide to have a drink at the campus pub. Following the next departmental seminar we both, without saying anything to each other, start walking to the pub to have a drink. At this point we can be said to be part of a joint commitment to have drinks together at the pub after the departmental seminar. If, after a third departmental seminar, I decide not to join you for drinks it seems reasonable that if I do not communicate this intention to you for you to accept, you have the right to rebuke me with “Where were you? We were supposed to go for drinks together.” In this case, we each owed it to the other to have drinks together; we were, as a “plural subject”, committed to do something together.

I have been claiming that this kind of account is foundational to our understanding

of institutions. In other words, if this practice of having drinks after departmental seminars carries on and involves, perhaps, an expanding and changing group, we might naturally call it an institution. It would be an informal institution, to be sure, but an institution none the less. However, I am not committed to this two-person, ephemeral group being an institution. The claim is not that all plural subjects are institutions. Rather, I am claiming that all institutions are plural subjects. The important difference between being a plural subject and an institution is the permanence of the plural subject.

Some have accused Gilbert of 'over-intellectualising' the notion of joint commitment and criticised her insistence that individuals must understand a joint commitment in order to be part of one (Baumann 2010, p. 15). They maintain that it is "psychologically implausible" for an individual to understand a joint commitment. Presumably, the thought is that the conceptual work that goes into understanding plural subjecthood and joint commitment is far too complicated for people to have intuitively. This criticism misses the mark because Gilbert is clear about the idea that individuals do not need to understand her technical treatment of joint commitment to understand the phenomenon implicitly. This does not seem to be an extraordinary claim. It is a case of 'knowing how', rather than a case of 'knowing that' (Ryle 1945). Just as in Heidegger's analysis of hammering in *Being and Time*, a hammer does not need to be known theoretically for us to be able to manipulate it (Heidegger 1962, pp. 69ff; cf. Carman 2003, p. 19n24). Joint commitments are something that is 'ready-to-hand' (*zuhanden*) and Gilbert is giving us a theory on how to understand them.

We can see this more clearly by looking at people's intuitive reactions to uses of 'we'. For example, if in line at the cinema a stranger were to turn to you and ask, "Should we see a film?" you would not assume that the 'we' was meant to include you. Most likely,



you would think that she is asking something like “Do you think it is a good idea if my friend and I see a film?” – a weird question, perhaps. In fact, it would be quite odd if this person was including you in this 'we'; at best, it would be what Gilbert calls an “initiatory” use of 'we' – an attempt to create a plural subject with you (Gilbert 1992, p. 178).

Now take a slightly different situation. You and a good friend decide to go to the cinema. Perhaps this is decided in the following dialogue: “Should we go to the cinema?” “Sure, let's go!” Now in line at the cinema, your friend turns to you and asks the same question as above, “What should we see?” In this situation, the 'we' clearly includes you. You and your friend have decided to go to the cinema together. Conversely, it would be odd if the friend turned to you and said, “I am going to see this film, what are you going to see?” Surely, you would respond, “I thought we were seeing a film together!”

Therefore, I take it that Gilbert's project as a whole is a promising quasi-transcendental proposal on how to make sense of institutions. She takes as a given that joint commitments do occur and she is trying to map out the conditions for the possibility of this phenomenon. By doing so, she thinks we can gain a better understanding of the social world. If this is the case, then the argument for being overly intellectual misses its mark; it is not the phenomenon that is overly complex but rather it is the theory which is forced to be complex in order to capture all the important aspects of the phenomena.

To understand the important use of the 'we' and plural subjects further, we need to understand plural subjects and establish their particular identities. Plural subjects are defined by and gain their identity from their constitutive commitments. For example, what makes baseball an institution is the constitutive commitments that make it up: the rules and roles of the game. If these rules and roles are not abided by then the activity, whatever it is, is not baseball. A constitutive commitment, as we have been seeing, comes about only

when each party expresses readiness to commit. There can also be derived commitment. These commitments can only occur when a plural subject is already formed.

It is these derivative commitments in which authority in our everyday sense can first be seen. That is to say, the individual who is appointed by the group to be *in* authority is now owed obedience – this is akin to Hobbes's 'Leviathan'.

An example is helpful. To use the paradigm case Gilbert often uses to illustrate joint commitments, we look at the case of 'walking together': two individuals, Leo and I, both express a readiness to walk back from a conference to a hotel together, either through a verbal or non-verbal exchange. In this case, we are not individually committed to walk back together but *committed together* to walk back to the hotel. Following this collective commitment, neither of us is able to rescind the commitment without the other's agreement. It is not that Leo is committed to go with me and I am committed to go with Leo, in the sense that Leo can rescind his commitment if he gets tired or wants to do something else. Rather, we are both committed as a body, or a 'plural subject', to walk back to the hotel together – neither individual can change this unilaterally. Furthermore, if I begin to walk faster than Leo, Leo is in a position to rebuke me for breaking the commitment *to walk together* and I may feel remorse, not for walking too fast but for breaking the commitment – for not giving what was owed to the group.

In the case of walking together, a derivative commitment can be formed where Leo is *in* authority. This can happen if, say, as we begin to walk I ask, "Which way is the hotel?" and Leo tells me it is this way. At the next street, Leo then turns left and I follow. At this point, it can be said that Leo is *in* authority when it comes to decisions about which direction to go. Leo now has the proper standing because I owe my deferral to him: we have formed a derivative joint commitment to follow Leo's directions.

Now, of course, if Leo starts to go the wrong direction and I become aware of this, I might have overriding reasons not to follow him. I might even object that he is going the wrong way. However, as we see in the next chapter, it does not follow from this that the reason from within the commitment completely vanishes. I am still committed to the group. However, I have overriding reasons not to conform.

The owing relationship that is characteristic of a group's constitutive commitment gives each member a directed obligation. Hence, a group commitment gives each member a sufficient reason to conform to the group's commitment. Further, it gives each member, qua member, the right to hold other members accountable. The members of the group then are “publicly bound (within the group) to maintain and satisfy the [constitutive commitments] and bear responsibility for these to matters, hence for the group members' acting correctly as group members” (Tuomela 2010, p. 16).<sup>\*</sup> This is where authority begins to emerge within a group commitment. If each member is bound and responsible to the constitutive commitment, then there is a sense in which members of a group commitment have authority over the other members of the group: each member has the authority, as a *representative* of the group, to hold all other group members to account for their conformity. It is, however, the constitutive commitments of the group itself which are the grounds of these authoritative relationships. It might be appropriate to call this a kind of “*constitutional* patriotism” (Habermas 1998 p. 500; cf. Gilbert 2014, Ch. 16) in which all members qua members are bound to the constitutional commitments. This is the internalising and socialising role that group membership plays in our practical lives.

<sup>\*</sup> Tuomela's original phrasing is in terms of 'ethos' of the group, which he defines as “the set of *constitutive* goals, values, beliefs, standards, norms, practices, and/or traditions that give the group motivating reasons for action.” My use of constitutive commitment as opposed to 'ethos' is merely terminological.

Now we should turn to an object we put off at the beginning of this discussion regarding the issue of large groups. A. John Simmons formulates this large group objection by telling the famous story of the women of Königsberg setting their clocks to Kant passing on his daily walks (Simmons 2001, pp. 76-77). Simmons rightly contends that the women of Königsberg have mere reasonable expectations, not entitlements, in regard to Kant walking by at the same time. The mere fact of regularity is not enough. A contrasting example that Simmons proposes that would entail obligation and entitlements is one of a group of friends who meet every Friday to play bridge. He further contends that the political “more closely resemble the indirect, impersonal relationship between Kant and the housewives [...] than they do the direct and personal relationship [...] between the bridge-playing friends” (ibid. p. 77). This is a significant objection if this understanding of joint commitment is going to help account for state authority. However, if we remember that, according to Gilbert, the only major requirement for a joint commitment is an intention to be committed which is *common knowledge* and that commitments form rather than *are* formed, we can get a sense of how this can occur with large groups.

According to Gilbert, all that needs to occur is that the inhabitants of an island start to refer to themselves using a collective ‘we’, perhaps in a newspaper that is easily available and read on the island. On more than one occasion she has pointed to Benedict Anderson's classic work, *Imagined Communities*, as a potential explanation of how this ‘we’ formation of large groups may come about (Gilbert 2000, p. 119<sup>n27</sup> and Gilbert 2006b p. 272<sup>n43</sup>).

Briefly, Anderson defines an imagined community in the following way: “It is *imagined* because the members of even the smallest nation will never know most of their fellow-members, meet them, or even hear of them, yet in the minds of each lives the image

of their communion” (Anderson 2006, p. 6) and it is “a community, because, regardless of the actual inequality and exploitation that may prevail in each, the nation is always conceived as a deep, horizontal comradeship”. (ibid. p. 7). This (imagined) comradeship is fostered by the print media (and now by electronic media) and a changed understanding of time in which events are only connected by “calendrical coincidence” (ibid. p. 33) and are brought together to create the imagined linkage. This new understanding of events has been fostered by the print media and the novel (Anderson 2006, p. 35; cf. Benjamin 1999). Whether this is sufficient for Gilbert's needs is too large a question to presume here, though it does seem like a live and promising option.

Furthermore, “existing members may establish rules that determine who may become a member” (Gilbert 1992, p. 233). In the case of our island, the initial forming group then can establish a rule that states 'only people born on our island are citizens once they reach the age of majority and are citizens by fiat.' A child growing up on this island would surely learn of the group's existence and that they are a part of it through standard socialisation processes.

Apart from this more practical issue of how there can be common knowledge of a joint commitment within a large group, we must also remember that, according to Gilbert's account, there is nothing that restricts coerced agreement, i.e. one can be coerced into being part of a joint commitment. How can this be? In her article, “Agreements, Coercion, and Obligation”, she criticises two arguments against coercive agreement: 'the Obligation Argument' and 'the Voluntariness Argument'.

The first argument, 'the Obligation Argument', amounts to saying, “A genuine agreement cannot be made in the face of coercion” (Gilbert 1996, p. 283). The idea is that agreements conceptually entail obligation. However, if one makes an agreement in the face

of coercion, then there is no obligation (ibid. pp. 282-283; cf. Gilbert 2006b, p. 77). However, it does not seem that it is at all a “contradiction in terms” to talk about coerced agreements (Gilbert 2006b, p. 78) - “He forced me to agree” is not a linguistic oddity (Gilbert 1996, p. 286).

The second argument, 'the Voluntariness Argument', claims that agreements by definition are voluntary and coercion does not allow for voluntariness. According to Gilbert, this argument exploits an “ambiguity in the notion of voluntariness” (ibid.; cf. Gilbert 2006b, p. 78). To help clear up this ambiguity, she introduces the idea of the 'decision-for' sense of voluntariness (ibid.). In her later work, this becomes the “intention” sense (Gilbert 2006b, p. 78). That is, an agreement is not voluntary if there has not been a decision in favour of the said agreement. For example, a person walks up to Bob Dylan and asks, “Can I have your autograph?” and Dylan without much thought, because this is a common occurrence for him, signs his name on the paper handed to him. If the paper was actually a recording contract (and there was no prior understanding of this, i.e. this person was a complete stranger) Dylan did not intend to sign a contract. He was duped. That is, this situation is not voluntary in the decision-for sense. Now, coercive agreements, on the other hand, are clearly voluntary in the decision-for sense – we know what we are doing when we make a coerced agreement.

Nor is it plausible to suppose that everyone becomes so unnerved in the face of coercion that they are incapable of making up their minds at all, that they are, so to speak, rendered witless (Gilbert 1996, p. 287)

In both of these cases, if these arguments are correct, it appears that one can be coerced into an agreement.

In Darwall's discussion of Gilbert, he objects to the idea that there can be coerced agreement. He does not argue from the standard presupposition of most theories of consent

that coerced agreements are non-binding, but rather the second-personal recognition underlying any genuine agreement rules it out:

the capacity of individuals to make agreements and form plural subjects depends upon their already presupposing one another's second-personal standing in seriously addressing each other in the first place. So any resulting obligation depends both on what they presuppose, namely, that they both have the requisite authority, and on their addressing one another on terms that presuppose this. It is the terms of this standing as mutually accountable persons in general that then gives them the authority to obligate themselves especially to one another through the terms of their agreement (Darwall 2006, p. 202).

Thus, without presupposing the 'equal dignity' (ibid. p. 202<sup>n35</sup>) of each party, the reasons to act are 'defeasible not just overridable' (ibid. p. 186<sup>n9</sup>). However, this seems premature on Darwall's part. As Honneth has clearly shown via Sartre's encounter with Frantz Fanon, there can be social situations that “distort intersubjective relationships of reciprocal recognition” and which “represent interactive relations that demand from both sides the simultaneous denial and maintenance of relationships of mutual recognition” (Honneth 1995, p. 157). In other words, Darwall is overloading his notion of 'equal dignity'.

It is possible, and this has occurred throughout history, that individuals convince themselves of being inferior to others and to respect their own dignity is to respect them in this way – think of an 'Uncle Tom'. This person would have an individual commitment, an internalised ought (Gaus 2011, §12), towards the other: “I ought to be...”. Paradigmatic instances of such distorted relationships of mutual respect are colonialism, racism and sexism. From our contemporary perspective, it is clear that these relationships are distorted. Hence, we find these type of relationships morally unjustifiable. We cannot understand the reasons that either side of the asymmetrical relationship give; they, as it were, speak a different moral language. However, from their own perspective, this was not entirely clear to the participants themselves, even with the “quasi-necrotic” behaviour they manifested

(Honneth 1995, p. 157). Darwall's criticism misses the mark because he confuses moral justification with the bindingness of standing. As we will see in the next chapter, there is no reason to think that joint commitments create moral obligations. All they seem to create is the direction of obligations between individuals in institutions. Furthermore, he takes a spectator's position outside the institution and looks for justification as such. However, this is something that can only be done given our knowledge of said biases. From the perspective of the participants, these biases remain hidden, which is what makes them dangerous and distorting (Peter 2009, p. 135).

### 3. The Authority of Us

How can all of this account for political authority? Now we should be able to understand why all authority is located with the members of the group as a whole. When a joint commitment is first formed, each member of the joint commitment is *in* authority in that “[e]ach, as a member of the whole, has the standing to demand compliance, issue rebukes for non-compliance, and the like” (Gilbert 2006b, p. 253). In other words, each member of a political society owes the community as a whole conformity to the commitment to govern together. If the group as a whole decides that a certain act is forbidden, all the members are in the appropriate position to demand compliance.

Furthermore, there can be a derivative commitment that each has expressed a readiness to accept the yearly election of a leader through a vote; the leader thus elected will have sole authority over issuing edicts and punishing non-compliance. It could also be the case that the group as a body expresses a willingness to commit to having a small group be the sole authority over issuing edicts and another small group having the sole authority to punish non-compliance. It is in this way that Gilbert helps to shed light on the idea of



popular sovereignty. As she points out,

one interesting aspect of this conception is that it helps to explain the sense one might have that the source of political authority in every case lies with 'the people'. Here, a joint commitment of the whole population in question – the people – is taken to underlie whichever kind of rule is in place (Gilbert 2006b, p. 213).

At this point, it is helpful to introduce the notion of a status function to make sense of this claim. A status function, according to Searle (1995), is when through an act of declaration we count one thing as something else. This is like the famous pillar on the Isthmus. By carving into the east side of the pillar that “Here is not Peloponnesus, but Ionia” it became the case that the east side of the pillar counted as Ionia (Plutarch 1960, p. 31; cf. Turner 2009). Or perhaps in a simpler example which Searle favours, that of money (Searle, 1995). With paper money, we all know that it only counts as valuable because we take it to be valuable; we make it the case by taking it to be the case. We place on these little bits of paper the status of money.

The same holds true for authority. It is not until the group commitment is formed that there is authority – authority is parasitic on the commitments to the group. Unlike consent tradition ideas of authority, the creation of a group commitment is not a transfer of authority but rather the creation of a new source of authority. The only time *A* has authority as a representative of the group is when *A* has been designated by the group as the authority: “They [the group] accord to him [the authority] a status, and with that status a function. He now *counts as* their leader” (Searle 2003, p. 201). Only as a group and not as individuals do they have this power, as it is a group that they have brought their representative authority figure into existence. This is not necessarily to rule out the moral authority of pre-group individuals. Rather it is to claim that there is another source of authority in the creation of a social group.

Jean Hampton has developed a model similar to the one being developed here. On Hampton's 'convention model', political authority is:

invented by those people through their participation in a governing convention, by which they give what I [Hampton] call convention consent to their regimes. Such consent is insufficient to morally legitimate the regime in full, but it forms the foundations for such legitimation [...] The invention of political authority involves creating authoritative offices such that when officeholders issue commands, they give the rest of the populace reasons to perform actions that preempt other reasons these people have to do other things (Hampton 1997, pp. 112-113).

However, one obvious difference between Hampton and Gilbert is the following. To be minimally legitimate, for Hampton, the authority must be “at least minimally rational and moral” which is internal to the creation of legitimate power (ibid. p. 112). This is captured in her discussion between a ruler acting as the people’s agent and a master who rules purely by force (ibid. pp. 86-94). Gilbert does not make such a distinction. It should by now be clear how standing is created in Gilbert's account, but what about justification of content? This is what we turn to in the next chapter.

## ***Six: Social Ontology and the Question of Content: Unity through Social Reason***

In the previous chapter, we developed an understanding of the background relationship which establishes the standing of authority. Standing, remember, is the normative relationship between authorities and their addressees. It was argued that the notion of a plural subject is a particularly powerful way of understanding this dimension of practical authority and the way it emerges between individuals when they form plural subjects which are the basis of institutions. We can characterise this relationship as one of owing. The basic idea is the following: when *B* owes conformity to *A*, then *A* has authority over *B*. The conformity that *B* owes to *A* is derived from the conformity *B* owes to the constitutive commitments of the institution of which *A* and *B* are both committed members.

To come to grips with the other aspect of authority which was delineated in Chapter One, viz. what justifies the content of authoritative commands, we must focus on the constitutive commitments of plural subjects themselves. Particularly, we must understand the relationship between constitutive commitments and the validity of particular commands.

To do so, we turn to the connection between institutions and the idea of social reasons which is the topic of Section 2. This connection provides an internal criterion for judging the validity of commands. That is, this internal connection answers the question of content. The basic idea is to show the symmetry between reasons which are constructed by individuals through the act of willing and social reasons which are constructed through the collective deliberation. This type of symmetry is not unique in the history of philosophy. In fact, it is similar to the symmetry proposed by Plato in *The Republic* between individual souls and the *polis* (Plato 1997c, 368c–369a and 434d–435a). In other words, what we

show is that institutions are able to be a source of social reasons just as the individual wills can be a source of reasons for individual agents.

How does this answer the question of content? The question of content was about who benefits from commands. That is, it is meant to justify why authorities can give particular commands. In Raz's service conception of authority, we saw one answer to this question. Authorities are justified in giving commands when those commands get subjects to conform better to reasons they already have. The account here has a similar structure in that authorities are justified in giving particular commands when they get their addressees to conform to reasons they already have.

However, unlike Raz's account, the reason that the addressees already have are a general reason to conform to the constitutive commitments of the institution of which they are a member. In other words, unlike Raz's and Enoch's accounts which suggest that we need to look for the reasons that each individual already has outside of the authority relationship, the argument developed here aims to show that we only need to look at the authority relationship itself to see whether a command is valid or not.

I want to suggest that these claims can be made sense of – and are not as counter-intuitive as many seem to think – by looking at a distinction Gilbert has repeatedly made since 2006, that:

*Standing*, incidentally, must be sharply distinguished from justification. One may have the standing to demand something of someone, yet not be justified in doing so, in the circumstances.” (Gilbert, 2014, p. 111)

Or more forcefully:

By virtue of the existence of the commitment, and that alone, the parties have rights against each other to actions that conform to the commitment. As a result, they have standing to demand such actions of each other and to rebuke each other not so action [...] this is not to say that their making such demands or issuing such rebukes is always *justified*, all things considered. To say that someone has standing to do

something means simply that he is in a position to do it. If someone lacks the standing to do it, the question whether he is justified in doing it does not arise. For he *cannot* do it. One who lacks the standing to make a certain demand or issue a certain rebuke can, of course, utter a purported rebuke or make a purported demand. He can speak in a rebuking or demanding tone. His target, meanwhile, may have little interest in this if it is possible to question his standing actually to rebuke or demand (Gilbert 2006b, p. 147; cf. pp. 45-46, 103-104, 190, 245-255).

In other words, by virtue of being in the right relation with another – i.e. being part of a plural subject – one has the proper standing to demand  $\phi$  from other members; however, they might not be morally *justified* in making that demand. This implies that, in some sense, a plural subject is, in Georg Simmel's words, “composed of beings who are at the same time inside and outside of them” (Simmel 1910, p. 384; cf. Gilbert 1992, Ch. IV and Gilbert 2003 p. 57).

As we will see, Gilbert means by ‘justified’ in the above passages that the command is justified from a point of view, particularly the moral point of view, which is outside the joint-commitment itself. To understand this distinction further, she introduces the idea that the obligations within a joint commitment are direct, as opposed to moral obligations which are not. This is the topic of Section 3.

If this is correct, it suggests far-reaching consequences for the lives of groups generally and for political institutions in particular. If there is a necessary connection between institutions and social reason then we are moved closer to seeing the necessity of democracy. Further, there emerges a connection between institutions and epistemic justifications of democracy, particularly pure epistemic proceduralism. If I am right, then we move towards a vindication of Hilary Putnam's claim that, “Democracy is not just a form of social life among other workable forms of social life; it is the precondition for the full application of intelligence to the solution of social problems” (Putnam 1990, p. 1671).

However, this consequence cannot be fully worked out here. What will be established is that the internal standards of an institution give grounds on which to judge the legitimacy of a command. It does not, however, follow that this makes the commands just or, as Gilbert puts it in the above quote, *justified*. What would make a command just would be a just institution. The criterion for a just institution, however, must be brought in from outside the institution itself. What we show instead is when a command is valid according to the internal criteria of the institution itself.

### 1. Constitutive Commitments

As we have seen, one of the most important questions for an agent is, who can speak for the whole? This question is important precisely because it is what brings unity to the potential disunity of competing desires. However, there is a difficult dis-analogy between institutions and individual agency in that institutions are composed of individual agents. This is problematic in the following way: if autonomy is a constitutive part of agency, then how can individual agents subordinate themselves to the commitments of the institution? Bear in mind that this is the classic dilemma we started with under the own-action condition. Thankfully, Rousseau shows us a way out: “Since no man has a natural authority over his fellow-man, and since force produces no right, conventions remain as the basis of all legitimate authority among men” (Rousseau 1997, p. 44). In other words, what is required is a convention to which all agents commit themselves. However, as we have seen, this commitment cannot be the individual commitment of the individuals but rather a joint-commitment to the constitutional essentials of an institution.

Due to the infamous problems in the realm of consent, however, this cannot be a matter of actual consent in the way consent is traditionally understood. We are not asking

the traditional question of consent, which is the idea that each individual consents to be ruled by a central authority. Rather, our question must be the following: how can individual agents produce an institution with unified commitments that all the agents can recognise?

A good place to begin is with a point that Velleman has made. An intention is the resolving of a deliberative question. As he says, if the question is not up to you, you have nothing to settle (Velleman 2000, p. 203). Therefore, the question 'what should we do?' must be settled by us; you cannot settle our deliberative questions unilaterally (unless you have the authority to do so).

From Velleman's point, we can draw some important conclusions. First, and entailed by what was just said, we cannot think of the types of commitment of institutions as individual commitments. This was the argument of Chapters Three and Four. In short, individuals are not entitled to change the commitments of an institution unilaterally. If they are not able to settle an issue on their own because it is not up to them to settle, then it surely follows from this that they are not able to rescind the commitment by themselves, either. This is a point upon which Margaret Gilbert has been particularly adamant in insisting.

We have used the term 'constitutive commitments' several times throughout this discussion but have not done much to understand what these are. It is time to give an account of such commitments and their importance for institutions.

Constitutive commitments are, to quote John R. Searle, what "create or define new forms of behavior" (Searle 1969, p. 33; cf. Rawls 1999a, p. 49). In other words, the content committed to by the members which constitute the identity of the plural subject. It is these basic constitutive commitments that form the heart of justifying the contention of authority. That is, what authorities can command is determined by what the constitutive commitments

of the group give them power to command. If someone in the role of authority commands something which they are not entitled to command by the constitutive commitments of the institution, then it is an invalid command. The authority does not have the power to command such things.

The claim of this section, therefore, is that there lies in all institutions a way to see if further commitments and, most significantly for our purposes, commands are justified in terms of more basic commitments. This is a fundamentally important way in which standing, being a product of a joint commitment, is related to the content of such a commitment.

The basic thought is this: the constitutive commitments of an institution establish limits to what the authorities embedded in the said institution can command, ensuring that the authority is not entitled to command anything that would be incompatible with the content of the institution's constitutive commitments. Like all members of an institution, the authority is also publicly committed to conform and respect the joint commitment which is at the heart of the institution.

This is a justificatory relationship between constitutive commitments and derivative commitments; the constitutive commitment has *internal* to it certain normative criteria according to which derivative commitments may be judged (MacIntyre 2007 and Marmor 2009 Ch. 2). One thing to note is that there is no reason why constitutive commitments cannot be nested. For example, within an institution a sub-group can be formed with its own constitutive commitments. The constitutive commitment of the sub-group is therefore a derivative commitment of the larger institution. In this way, the sub-group is constrained by the constitutive commitments of larger institutions as well as its own.

Indeed, we can see this kind of structure in the history of the United States. After the



American Revolution, a new constitutive commitment was formed. The identity of the new institution was determined by the commitment to this new constitutive commitment, viz. those committed to the United States Constitution. Within the framework of the Constitution, the process of judicial review has been established. Here we can see a primary example of how derivative commitments are challenged on the basis of the constitutive commitments, in this case the Constitution of the United States.

The idea of judicial review is that the derivative commitments (i.e. laws) that are created by the legislature – one aspect of political authority or “authorizing members” (List and Pettit 2011, p. 35) in the United States – can be reviewed by the judiciary to ensure that they do not violate the constitutive commitments of the United States (i.e. the US Constitution). One particularly appealing way to conceive of judicial review is as “a kind of rational and shared pre-commitment among free and equal sovereign citizens at the level of constitutional choice [...] limit[ing] the range of legislative options open to themselves or their representative in the future” (Freeman 1990a, p. 353). What is important here is not the contractarian element but rather the view that judicial review can be seen as a 'shared pre-commitment' which functions to eliminate legislative options. There are many complications and debates about judicial review and its legitimacy for democracy. We can, for our purposes, leave these aside. What is important for us is to see that if we take seriously the constitutional commitments that are fundamental to a particular institution as a 'shared pre-commitment', we can see the limiting effect they have on future decisions. These shared pre-commitments rule out certain decisions that authorities can make for the group and remain valid.

Roberto Unger has argued that contemporary jurisprudence has a “discomfort with democracy.” This discomfort “shows up in every area of contemporary legal culture...”

which has led to the “[f]ear and loathing of the people” which “threaten[s] to become the ruling passions of this legal culture” (Unger 1996, pp. 72-73). Jeremy Waldron agrees and has tried to rise to Unger's challenge (Waldron 1999, pp. 8-10). Something similar can be said about at least a portion of contemporary political philosophy when it comes to the justification of authoritative content.

For example, if we turn back to Raz's Normal Justification thesis (NJT) or Jonathan Quong's recent 'anti-perfectionist' or 'political-liberal' reworking of NJT as a duty-based account (Quong 2011, Ch. 4 §§4-5), we can see this trend. Although there are many fundamental differences between Raz and Quong, both of their accounts rely on the idea that what justifies the authority of the state and its commands is showing that we have “most reason to do what the state commands” (ibid. p. 111) In this way both are (or are potentially) anti-democratic in that the state has to be justified based on an independent criterion. For Raz, it is the ability to get us to conform better to reasons we already have. For Quong, it is the ability to get us to conform better to our natural duty to support just institutions reasonably. Neither appeal to the internal justification of the institutions themselves which have been created by individuals through forming joint-commitments; that is, the justificatory power of the popular sovereignty that we saw underlines the emergence of all institutions.

The implicit worry of those like Raz or Quong about democracy seems to come down to democracy coming to the 'wrong' answers; that it does not conclude in policies that get agents to comply better with reasons they already have (Raz) or decisions that are not reasonably just (Quong). If democracies do not come to decisions that meet this sort of external criterion then it does not issue authoritative commands and citizens have no reason to obey. This is tying justice and external justification too closely to legitimacy of

command.

Gerald Gaus, who is interested in 'public reason' and 'public justification', is also sceptical about democracy. Gaus's deliberative model, for example, relies on “Members of the Public” who are “idealized counterparts of actual members of the public, but they are not so idealized that their reasoning is inaccessible to their real-world counter-parts” (Gaus 2011, p. 276). Gaus wants to avoid appeal to *actual* people and *actual* deliberation. This is because Gaus thinks democrats are captivated by a romantic and “the highly ideal picture in our mind's eye of the Athenian polis” which can “only lead to authoritarianism and oppression” (ibid. p. 387) and that what we really need to do is take seriously our evaluative differences and concentrate on what is justifiable to “the Members of the Public” as such, “[d]emocratic procedures simply are not up to the task of collective commensuration” (ibid. p. 388).

Projects like those of Raz, Quong and Gaus are important and can help to give us a perspective outside the democratic process. However, in the following, I take the view that “no amount of insight into what might be owed to people will settle the question of how decisions about [policies] ought to be made” (Peter 2009, p. 1). Focusing on and trying to determine the justificatory power of procedures is important for real world politics as most agree, or popular rhetoric would lead us to believe, that it is commonly accepted that democratic governance is the only justifiable form of governance which we know of today. Most decisions are claimed to be democratic and there is outrage from many when governmental decisions are made behind closed doors. This can be seen, for example, in the growing concerns over the current secretive negotiations of the Trans-Pacific Partnership between 12 Pacific nations (the US, Japan, Australia, Peru, Malaysia, Vietnam, New Zealand, Chile, Singapore, Canada, Mexico and Brunei Darussalam). These negotiations

have been deemed secret by the Obama administration who are even denying members of Congress access to the terms of the negotiations and the Australian Government has banned reporters from being briefed on the negotiations.

What follows from such a project, however, is not an apology for present policies or the status quo. Rather, there is a critical edge viz. by determining the justificatory force of democracy, one can then see how well existing democracy compares to the ideal type that is developed in theory (Young 2002, pp. 10-11).

The relationship between standing and content, however, does not have to be limiting. There is also a positive side to how commitments can justify the content of commands. This is the case of the creation of political authority. When an individual or group comes to hold standing as a practical authority, it becomes a matter of status conferral: “They [the institution] accord to him [the practical authority] a status, and with that status a function. He now *counts as* their leader” (Searle 2003, p. 201). With this status function, the practical authority now has the deontic power to make a certain set of decisions for the group as a whole.

Let us now see how the conferral of a status function entitles a practical authority to give certain types of command – that is, how it is justified in giving commands of a certain nature. It seems natural to say that whenever someone is made the leader of a particular institution, they are made leader to use their particular deontic powers, viz. powers to change the normative circumstances of those she has authority over, for a particular set of purposes. It is these powers that determine what the authority is entitled to command or the content of the directives the authority is entitled to give.

In the political case, take again the United States. To simplify, Congress is entitled to make laws that change the citizen's normative circumstances. As we saw above, this

entitlement also has its limits, viz. commands must not be incompatible with the constitutional commitments of the institution. Even so, there is a great range of content-full directives that Congress has the right to give, such as to set federal tax levels and federal environmental protection laws.

Perhaps it would be helpful to give a simplified example. Let us take three people who have formed a joint commitment to read a particular book together. The group has also decided that one of its members, *A*, will lead the reading group. They accord *A* with the status function of leader. The group as a whole has given *A* the deontic power to make decisions for the group: *A* gets to decide when the group will meet, how to run the meetings, how much to read for each meeting, etc. When *A* makes a decision about these issues, it changes what the other members of the group ought to do. However, there are also clearly things that *A* cannot direct the group to do even though *A* has the standing to make decisions. For example, although *A* has the standing to dictate how much to read, it is not in *A*'s power to change the group's reading material, nor can *A* decide to spend the entire meeting without discussing the book despite having the standing to dictate how the meetings will run. In both these cases, the members of the reading group would have reason to protest because *A* is commanding things that violate the very commitments which are the basis of *A*'s authority in the first place.

Further, as a member of the group, *A* is also committed to the constitutional commitment of reading *this* particular book and discussing it with these people. *A* has the authority to determine a wide range of the group's activities but the content of *A*'s commands must still conform to the constitutive commitments of the group. This is due to *A* necessarily being part of the group commitment as well as being the leader of the group.

We can see a further positive power authorities have if we turn back to the

discussion above about how the background constitutive commitments of an institution are not necessarily formed through consensus. That is, there can be divergent understandings about what those commitments mean. What authorities are able to do is fix for a time the interpretation of these background agreements.

This is one of the fundamental roles of authority. There may be deep interpretive disagreements amongst members of an institution about what the constitutive commitments are that make up the background commitments. Authorities can help alleviate this situation in the following way: the authority can make public an authoritative interpretation of the background agreement that all members of the institution are accountable for following until the institution as a whole changes its mind.

To carry on with the reading group example, there are many different ways in which a reading group can be understood and structured. Two possible ways might be for there to be just a free-floating discussion when the group meets or there could be a more structured conversation where one member begins by providing a summary and questions for the group to discuss. When our reading group was formed, the members did not have a consensus about how the group would proceed during their meetings. Once there is a leader of the group, that leader can decide for the group on which of these two, or some possible other, ways of structuring the meetings. This fixes and makes public to all the other members how the group is going to understand what it means to have a meeting. They are all committed as group members to running the meeting this way.

Now, authorities will not resolve the disagreement but only alleviate some of the problems of disagreement when it actually comes to enacting policy choices. If there is deep disagreement between members on the appropriate amount of reading, *A* has the ability to determine how much the group will read per week. The other members may still

disagree, but they are responsible for following *A*'s directive on this issue until the directive has changed.

In this section, I have outlined some of the ways that constitutive commitments and authorities are important for our normative landscape. Further, we have begun to see how authorities can fix the interpretation of background commitments as well as determine further commitments of the group to solve issues of disagreement. What the authorities are doing when they change the normative landscape of the group is creating social reasons for action. Now it is time to turn to more detail about the reasons that the authorities are able to give and how these reasons come about.

## 2. What are Social Reasons?

What are social reasons? In this section, I argue that there is a symmetry between the reasons that are binding for agents that are the product of their practical deliberation and practical identities, and those that are binding for members of an institution which are the product of institutional deliberation and practical identity. The way that practical identities give reasons for action has been helpfully elaborated by Christine Korsgaard. An example she uses is that of a student (cf. Korsgaard 1996a, §3.3.4). The idea is that a student might be required to take a particular class, e.g. logic. Since the student here endorses practical identity as a student, this is a reason for her to act and her acting on this reason does not interfere with her autonomy. This is because she has endorsed the practical identity of being a student and this is what it means to be a student, to take the classes required of you. The reasons that are the product of institutional deliberation are what I will call 'social reasons'. They are social reasons because they are binding on all members qua members of a particular institution.

To see how this works with institutions as whole, rather than with individual practical identities, let us turn to a particular case. Presumably, any philosophy department is going to be constituted by several commitments. These commitments might be, in no particular order, the following: providing a quality philosophy education to undergraduates, a quality research environment for staff, support for graduate students, etc. All of these provide reasons for the department to act. However, as we know, philosophers tend to be individuals with strong opinions about how these commitments should be fulfilled. The politics internal to the department could be complex and contentious with deep disagreement as to what constitutes a quality philosophy education. Some may think that undergraduates should have a firm foundation in the history of philosophy while others may think that the best way to teach philosophy is through contemporary debates about current problems. How is the institution as a whole to make a decision on these types of issue?

Like the individual case where an individual deliberates about her practical identity, institutions also deliberate about their identities. However, rather than the individual deliberating alone, in the case of institutions it is a social form of deliberation. When an institution finds itself pulled apart by disagreement, the institution is no longer able to act and is forced to deliberate until the issues it is dealing with are settled, at least for a time. The issue of how to teach undergraduates must be settled in part for the philosophy department to teach successfully. In other words, we need to work out how the head of Reason in Plato's analogy (see Chapter 3.3) can be understood as the deliberation of the institution.

The proposal here is that the notion of social reason is helpful for this purpose. The term 'social reason', sometimes called 'public reason', has been highly contested and there



are many interpretations of how it is best understood. The general idea which I put forth here is that the norms structuring the life of an institution must be deemed acceptable by all its members. The general interpretative strategy I adopt here is the following: social reason is best understood in procedural terms, rather than in substantive terms. This amounts to saying that social reason is connected to the justification of the norms of deliberative processes themselves, rather than to what kinds of reason individuals can appeal to as deliberators (Peter 2009, Ch. 6; cf. Laden 2000, p. 551).

Rawls has referred to this as ‘constitutional essentials’, which he tells us are “what is of greatest urgency” for consensus (Rawls 2001, §9.3). Thus, the deliberative process itself must be justifiable in terms that all members can freely recognise, and any decision resulting from this process is legitimate if we remember that the content of a decision is legitimate only as the product of a deliberative process.

An alternative interpretation of social reasons is a substantive one. On this interpretation, all decisions must be justifiable, as well as accessible, to all via social reasons. The main problem with this interpretation lies in the difficulty in seeing the value of an appeal to social reason as a standard for all decisions. Social reason on the substantive interpretation is indeterminate, i.e. unable to generate a conclusive result. For the sake of illustration we may look at cases of substantial disagreement, as the very point of appealing to social reason is to deal with disagreement and reasonable pluralism.

This is the problem we encountered at the end of Chapter Two. Remember there we discussed Sandel's view on the distinction between fines and fees. From this, there was the claim that we should reject carbon trading markets because it fails to maintain the moral stigma of using more carbon than one should. There seems here to be a reasonable disagreement between Sandel and proponents of carbon trading markets. How can the

substantive form of social reasoning alleviate this agreement? More precisely, to what can the institution appeal to resolve the disagreement?

We now have two potential ways of choosing institutional commitments. The first option requires agreement on the *procedural* requirements of deliberation, whereas the second demands substantive agreement on every decision in the sense that, for a norm to pertain to the entire institution, it must be accepted by all participants. However, as with the Sandel case, it looks like he would not accept a norm of carbon trading. He would reject such a norm because it seems to keep the moral stigma he finds necessary when it comes to over-using common goods. Sandel does not seem to be unreasonable in rejecting this norm. It is a reasonable position to maintain, and to reject it as unreasonable risks undermining Sandel's status as a self-originating source of valid claims. This leads to a policy stalemate. Both sides see the importance of doing something for the good of the environment but deeply disagree as to how to proceed. Any institution caught in such a problem will be unable to act or will violate its members's status as a source of valid claims.

If this is correct, the second option is indeterminate with regards to choosing norms of action for the institution. This leaves us with the procedural option, which gives a way for an institution to make a decision, at least provisionally, that all the members are bound to accept as the norm of the institution. We find consonance between this and the following remark by Korsgaard:

In order to act together – to make laws and policies, apply them, enforce them – in a way that represents, not some of us imposing our private wills on others, but all of us acting together from a collective general will – we must have certain procedures that make collective decision and action possible, and normatively speaking, we must stand by their actual results (Korsgaard 2008, p. 247).

If this is not accomplished because of internal disagreement then deliberation as to what the institution ought to do will not cease. There will be continual conflict and disunity.

We can see this if we take a small group, say, a committee planning a party. If the committee does not jointly commit to the manner in which decisions will be made within the group, the group will be in continual disunity. Say there is a disagreement about what kind of food to buy for the party. It would be problematic if the group, working on a limited budget, were just to buy whatever anyone suggested. There must be a single decision made for the whole of the group. Let us say that the group recognises a simple majority rule decision procedure, in which case the individual members are obligated to follow whatever the outcome of the vote may be. This is how unity is brought to the group.

We can now see how important is the connection between social reason and institutions. The obvious reason for a connection, and a natural starting point, is that for an institution to be effective – just as in the case of individual agents – there must be a unity that overcomes the diversity and disunity internal to any institution. There must be something that holds the institution together as an institution in order for it to be able to do anything at all. It seems that for this to be possible, there needs to be some reliance on the idea of social reason which allows us to solve the disunity problem by giving us agreed-upon procedures to deal with disagreement and pluralism.

In the end, social reason is important to the constitution of group agency precisely because it is what determines the unity of the agent, which allows it to be effective. Social reason, then, plays the role of Reason in Plato's analogy for the group as a whole. As suggested above, it also gives us a criterion for the success of an action. When an institution acts for reasons which are not the products of the procedure agreed upon via social reason, this results in disunity within the agent itself. We can say, with Rousseau, that in this case the institution fails to have a general will but instead is determined by the private interests of particular wills (Rousseau 1997, Ch. 2.3.2). However, when an institution acts for

reasons which are grounded in publicly accessible norms of fair procedures then we can say that the group acted as a whole, as a general will.

We have seen how agents are normatively constituted and how there is a symmetry between the individual constitution of agents and the constitution of group agency. This has been the initial stage setting in order to answer the question of content. Let us now look more closely at how proceduralism helps us in answering the question of content.

### 3. Proceduralism and Content

The current debate about deliberative democracy has paid particularly close attention to the power of proceduralism. It is helpful to canvas this debate in order to understand more fully how proceduralism can justify the constitutional essentials of an institution and how this justification is transferred to the particular commands of authorities which, as we have seen, are roles that emerge from the workings of institutions.

Bernard Manin gives what can be seen as the shibboleth of deliberative democracy: “the source of legitimacy is not the predetermined will of individuals, but rather the process of its formation, that is, deliberation itself” (Manin 1987, pp. 351-352). More recently, James Bohman has strongly echoed Manin when he said that:

political decision making is legitimate insofar as its policies are produced in a process of public discussion and debate in which citizens and their representatives, going beyond mere self-interest and the limited points of view, reflect on the general interests or on their common goods (Bohman 1996, pp. 4-5).

Finally, to help clarify to what deliberative democrats see themselves as committed, David Held provides a particularly helpful summary of this model of democracy in this way:

The terms and conditions of political association proceed through the free and reasoned assent of its citizens. The 'mutual justifiability' of political decisions is the legitimate basis for seeking solutions to collective problems (Held 2006, p. 253; cf Habermas 1996, p. 448 and Cohen 1997, pp. 72-73).

Notice that these characterisations tend to rely on some form of idealisation, similar to the idealisation we saw in Darwall and Gaus's accounts. Iris Marion Young has presented a much more realistic view of the process in the opening pages of her *Inclusion and Democracy*. She presents a picture of a complex and slow process of democratic change. However, she agrees that the fundamental point of deliberative democracy is that “few question the legitimacy of the outcome because the process was relatively public, inclusive, and procedurally regular” (Young 2000, p. 3). She goes on to claim that “[t]he normative legitimacy of a democratic decision depends on the degree to which those affected by it have been included in the decision-making process and have had the opportunity to influence the outcomes” (ibid. pp. 5-6).

The important aspect of these views is that they claim that the justification of decisions is gained through the *process* in which the decisions are made. The decision made through the process, whatever it is, creates for the group social reasons for actions. If the members of an institution, who have recognised the decision-making procedure as a constitutional commitment of the institution, do not conform their actions to the social reason created by the decision procedure, then they are violating their commitment to the institution. Furthermore, it looks as if the institution has the right to hold accountable and rebuke those who do not conform.

Now we are able to understand the underlining structure of this situation more fully. The members of the institution have expressed readiness to recognise a certain decision procedure. That is, they have formed a joint-commitment to conform to any decision produced through the procedure of their institution. The decision procedure is analogous to the deliberation which individuals must go through when they step back from their desires and reflect on what to do. To be able to act again, the individual must endorse which ends

to pursue. The only difference is that the will of the institution is the pooling of many wills to form a general will and what this will create is not individual reasons for action but social reasons for action. Social reasons are binding on all members of the group qua being the result of their pooling of wills.

The process element is the element that most strongly sets deliberative democracy apart from aggregative conceptions of democracy. In aggregative democracy, decisions gain their legitimacy through a synchronic voting procedure; deliberative democracy claims that legitimacy is garnered through a diachronic deliberative procedure, a process that is slow, fallible and complex. It is important to point out, however, what is really at issue. The issue that is at stake is the very identity of the institution as such. When members of an institution are deliberating what to do, they are also deliberating what kind of institution they are going to be going forward, they are deliberating the very identity of the institution. The reasons this is so are because they are deliberating how best to interpret the constitutional commitments that are the basis of the institution's life.

There is, as Fabienne Peter has argued, “some sort of an agreement about the two main features” of deliberative democracy: “(1) decision making is based on *public reasoning* (2) under conditions of *political equality or fairness*” (Peter 2009, p. 31). In order to understand the legitimating virtue of deliberation, it is imperative to understand these two features. Let us look briefly at each feature to see how it factors into the justificatory task which deliberative democrats support.

First, public reasoning, or what I have been calling ‘deliberation’, is the giving and taking of *reasons* for or against a particular policy in the public sphere. Gutmann and Thompson put the point succinctly when they say that the difference between deliberative and aggregative conceptions of democracy is that the deliberative conception “asks for

justifications” (Gutmann and Thompson 2004, p. 13). The public sphere and the condition of publicity is important because it necessitates that reasons given in the public sphere, which is the sphere which incorporates all the members of a particular institution, must be potentially acceptable by others. This excludes – not formally, but pragmatically – the individual reasons that the members of the group have qua individuals from doing the justificatory work. That is, a claim of individual preference will do little to convince others that a particular policy is really the best policy to adopt.

Gerald Gaus has raised a worry about the reasonableness of subjects which can also be interpreted as a challenge to the idea that deliberation has epistemic value. He recounts the belief perseverance experiments of Lee Ross. The experiments showed that when subjects were induced to hold an unjustified belief and

[t]he subjects were above normal in intelligence, and their beliefs were subject to far more rigorous criticism than are most of our political beliefs. In spite of all this, subjects continued to hold beliefs that seem manifestly unjustified (Gaus 1997, p. 213)

The worry here seems to be, how can the exchange of reasons and criticism be any justification for the beliefs and actions of an institution if individual beliefs are so recalcitrant to rigorous criticism?

However, there is another aspect to deliberation that is of equal importance that Gaus's worry does not touch. This is the element which Young calls deliberative democracy's 'disclosive' effect. That is to say, the point of deliberation is at least partly to open up new possibilities that individuals may not have factored into their preferences. Contrary to Habermas's overly rationalistic theory, deliberation does not necessarily lead to changes in preferences by appeal to the force of the better reason.\* Rather, deliberation can

\* The 'disclosive' power of deliberation is often overlooked by Habermas, even in his own writings (cf. Kompridis 2011 and Miller 2011).

bring forth new options and directions for citizens to factor into their preference sets.

This transformative effect is essential to the epistemological justification of democracy and the legitimating force of deliberation (Putnam 1990, p. 1671). Putnam summarises the thought in the following way:

The fact that someone feels satisfied with a situation means little if the person has no information or false information concerning either her own capacities or the existence of available alternatives to her present way of life. The real test is not what women who have never heard of feminism say about their situation; indeed, it is hard to see how the situation of a chauvinist woman in India is different from the situation of a chauvinist woman in this country thirty years ago who had never been exposed to feminist ideas. Such women might well have answered a questionnaire by saying that they were satisfied with their lives; but after realizing the falsity of the beliefs on which the acceptance of their lives had been based, the same women not only felt dissatisfied with those lives, but they sometimes felt ashamed of themselves for having allowed such a belief system to be imposed upon them (ibid).

The 'disclosive' effect of discourse, or rather the revealing of new options and possibilities through disagreement, and the procedural adjudication of disagreement, has been “a constant engine of change” (Hampshire 1989, p. 55). These disagreements and the 'disclosive' effect they help to produce are important because, as Stuart Hampshire tells us:

Every person and every social group is to a greater or lesser extent blind to many of the injustices of its time, because its own culture and education, supporting a particular way of life, represents embedded and distinctive features of this way of life as unavoidable features of human life in general (ibid. p. 59)

This brings us back to the discussion of how social situations can be distorted that we saw in the previous chapter. Notice, however, that what drives the social innovation and change is not throwing up one's hands and saying that the bindingness of institutions do not hold because they are, from a backwards-looking glance, unjustifiable. Rather, the authentic change in the lives of institutions comes from the disagreements internal to it and forces a change in the very identity of the institution. The thought might be that we are not looking



for revolution but conversion and rebirth.\*

This is exactly the point of Putnam's example of the possibility of a less chauvinistic culture. However, the importance of this effect does not have to be so grand but can be something as simple as the exposure of a lesser-known candidate. Fabienne Peter's example illustrates the middle ground between the reason-based and the 'disclosive' aspect of deliberation:

As a result of deliberation, some initially expressed preferences will seem unsustainable and be transformed, and new preferences will emerge during the process of public deliberation. Deliberation may, for example, strengthen the reasons people have for not endorsing a policy that discriminates against women. Or it may lead some to abandon their strong initial preference to elect a woman to a high political office, if the deliberative process generates strong reasons in support of a previously lesser known male candidate. (Peter 2009, p. 33)

There is an important point here that is crucial in the following. Just as in scientific experiments, hypotheses are tested on how well they work and revised accordingly (of course, this is an oversimplification but the point is the emphasis of the back and forth movement between hypothesis and critique based on 'success' or 'failure' of experiments), the same is necessary in the public sphere. Hypotheses, i.e. policies, are to be proposed and, if implemented, criticised based on their success or failure. This may seem common sense but this is what is overlooked by some when thinking about the transformative effect of deliberation. Part of the deliberative process is *showing* that certain policy choices are not working. Of course, recalcitrance and perseverance of beliefs is a problem but it is no more of a problem than in any other domain of inquiry – deliberative politics is a long, slow process that makes mistakes and gets things wrong, but why should it not be and why would this be a criticism of it? John Dewey, in Putnam's reading, brings these two elements together in what he calls, following C.S. Peirce, 'the scientific method': "the scientific

\* Compare with Stanley Cavell's remarks about education for grown-ups in his *The Claim of Reason* (Cavell 1999, p. 125)

method is simply the method of experiment inquiry combined with free and full discussion – which means, in the case of social problems, the maximum use of the capacities of citizens for proposing courses of action, for testing them, and for evaluating the results” (Putman 1990, p. 1671).

This leads to the main point of Gaus's criticism and also indicates the way out of it. What he has called, and rightly criticised, the ideal of social reason in a “radical sense” is true in that “only reasons that can be embraced by all of us are truly public, and hence justificatory” (Gaus 1997, p. 205). There certainly is no reason to believe that social reason will ever have the outcome of “wide, though of course not complete, actual consensus on political outcomes” (ibid. p. 206; cf. Gaus 2011, pp. 387-388). There will, as long as we have a deeply pluralistic society, be disagreement about justice and the common good.

Gaus's criticism is directed at deliberative democrats like Cohen and Benhabib who believe that aim of deliberation is

to arrive at a rationally motivated *consensus* – to find reasons that are persuasive to all who are committed to acting on the results of a free and reasoned assessment of alternatives by equals (Cohen 1997, p. 75; Benhabib 1994, p. 31).

The idea that these deliberative democrats are working with is one in which public deliberation should lead to substantive consensus on the issue before us. It seems that Gaus's point is well made against this type of theory.

However, not all deliberative democrats agree that we should focus on consensus and the common good, e.g. Iris Marion Young (2000) and, following her, Fabienne Peter (2009). For Young and Peter argue rather that 'the politics of difference' can be a resource for democracy which, through the situatedness of different individuals and groups, can, because of

the plurality of perspectives they offer to the public helps to disclose the reality and

objectivity of the world in which they dwell together [...] By communicating to one another their differing perspectives on the social world in which they dwell together, they collectively constitute an enlarged understanding of the world (Young 2000, p. 112; cf. Peter 2009 p. 35f).

This gives us a basic understanding of what it means for decision-making to be based on 'public reasoning.' It is not that the public needs to form a consensus on any particular topic but rather that the reasoning that goes into decision-making must be out in the open. The importance of this, as we have seen, is more to do with bringing forth new possibilities and perspectives on how an institution should think of itself.

There were two claims made above that may seem contradictory and hard to square. These were: (1) the rejection that deliberative democracy needs be about a search for a consensus and (2) that there is an epistemic dimension (the transformative effect) to deliberation. If deliberation is not a search for a consensus then what is it searching for – if anything - and what is the importance of the epistemic dimension? There are two answers which are currently in the literature: a rational proceduralist answer and a pure proceduralist answer. I take each in turn.

One understanding of epistemic proceduralism is David Estlund's *rational* epistemic proceduralism. Estlund rejects pure proceduralism, advocating that it must be supplemented by procedural impartiality, which has a *tendency* to 'track the truth.' Estlund's idea is an updated version of Rousseau's general-will, where citizens ought to follow the *public view* in so far as its procedures reliably generate results that are better than random (Estlund 1997, p. 196-198). There is a degree of truth to this; however, Estlund inflates the importance of the tendency to 'track the truth'. This inflation is due to Estlund's evaluation of pure proceduralism as equivalent to the "flip of a coin" because it lacks a "cognitive process" (ibid. p. 176-196; cf. Estlund 2008, p. 82). This might be true

for aggregative models of pure proceduralism, however, as Fabienne Peter notes, Estlund's account minimises deliberation's constructive power.

According to Peter, Estlund fails to acknowledge that "the epistemic dimension may be rooted in a fair decision-making process" (Peter 2009, p. 82). In other words, *contra* Estlund, fair procedures, when deliberative, have a cognitive element (as discussed above), which gives one reasons to *doubt* that one's "own moral judgement about the outcome is supreme" (Estlund 1997, p. 195; cf. Estlund 2008, p. 108). That is, he does not see why one would change one's mind based on the *outcome* of deliberation alone. Estlund in this passage is making the point that, after deliberation, one's own deliberation is superior to that of deliberation in general. However, it does not follow that because one does not change one's mind after deliberation that one does not have reasons for one's confidence to diminish (Peter 2012). Therefore, Estlund's worry is placated once we recognise fair deliberative procedure's "knowledge-producing potential" (Peter 2007, p. 343). There is, then, no need to qualify epistemic proceduralism with the procedure-independent criterion of 'better than random'.

There seems to be good grounds to accept the pure proceduralist account of democratic authority in regard to practical cases. Furthermore, most decisions that an institution needs to make are practical or come down to a practical issue such as which institutions ought to receive resources at the expense of others, e.g. should we build a new school or a new police station? This is not just an issue of expedience or efficiency but an issue of what weights we should give to different ends. It may even be the case that many issues which appear superficially to be theoretical will, in fact, be practical. For example, is the teaching of intelligent design allowed in public science classes? Many see this debate as being about the truth or falsehood of intelligent design. Yet the issue really is about freedom

of thought or conscience and how it ought to impact (or ought not to impact) the education of children in a free society. The truth of intelligent design is a different question to whether it should be taught or not in publicly funded schools. According to the deliberative democrat who accepts pure proceduralism, whatever the answers to these questions are, what is justified to enforce, i.e. what an institution has authority to do, is a matter of what falls out of the deliberative process itself.

The second feature of deliberation democrats share is the idea of political equality, fairness or inclusion which amounts to having the *opportunity* to participate in the political public sphere. As we saw above, Young takes inclusion and equality to be fundamental to the legitimacy of democratic decisions:

a democratic decision is normatively legitimate only if all those affected by it are included in the process of discussion and decision-making... Not only should all those affected be nominally included in decision-making, but they should be included on equal terms (ibid. 23).

This runs together her comments on 'inclusion' and 'political equality' but, as she says herself after this passage,

While I have distinguished the terms 'inclusion' and 'political equality' in order to specify their normative import, for the rest of this book when I refer to a norm of inclusion I shall understand it to entail the norm of political equality. (ibid. pp. 23-24)

It is best to understand this feature of political equality as equal opportunity to participation in a broad sense and not limited to formal argumentation.

Bringing these two features together, we can give a statement as to how deliberative democrats see the legitimization of policies: through the procedure that is characterised by an exchange of reasons for and against constitutive commitments, all who are affected have the opportunity to participate provisionally which legitimises the outcome until the institution as a whole is convinced to reconsider. This is what makes the constitutive

commitments of the group maintain their popular sovereignty and gives justification to the constitutive commitments. All members of the group as group members maintain the right to question the content of the joint-commitment and the institution of which they are part. The epistemic benefit of doing so is to disclose new options for how the group can identify itself.

Whenever these constitutive commitments are settled, the further content of the institution's decision making is bound to conform to it. As long as the institution's further commitments are within the limits of the constitutive ones, they gain their legitimacy and force from the basic commitments.

There is a worry that follows from this type of position which has been indicated several times. Namely, it looks like the consequence of such a proceduralist interpretation of what justifies constitutive commitments of an institution will lead individuals to be bound to an institution which gives them social reasons to act which are objectionable. Not only that but that they are obligated to follow these social reasons. Furthermore, how is one to criticise one's institution if one is bound to these social reasons? Are they not, after all, the individuals' reasons? This is the major theme of the concluding section.

#### 4. Directed Obligations: Social Standpoints and the Tragedy of the Social

One of the major obstacles to an adequate understanding of the bindingness and obligations that are part of institutions is the type of obligation which is at stake. The argument in this section is that the type of obligation which institutions generate are of a different kind from other types of obligation, particularly moral obligations. There are two consequences of this fact. First, understanding these different types of obligation allows us to understand how conflicting obligations and standpoints enable criticism of institutions.

Second, these conflicts are irreducible and tragic in the sense that one has normative reasons to regret not being able to fulfil both conflicting obligations.

Many simply see obligations and the reasons for action which they generate in an univocal way. Accounts which rely on the univocity of normative reasons, e.g. that all normative (practical) reasons are moral reasons, hide or make it impossible to make sense of the tragic character of the human condition. Typically these positions see all normative reasons as being straightforwardly weighable against each other (Scanlon 1998, Raz 2011 and Parfit 2011).<sup>\*</sup> On these univocity accounts, the problem of what we have most reason to do is an epistemic problem rather than a metaphysical one, viz. the problem is about our limited abilities as practical agents to determine what we have most reason to do. It is incompatible with this view that there is still conflict which an agent ought to regret for not being able to fulfil both conflicting reasons.

A non-univocal account of normative reasons, on the other hand, argues that there can be conflicts between what we morally ought to do and what we ought to do from the perspective of other practical identities (Korsgaard 1996b). This is a deeper problem than our epistemic limitations. No amount of knowledge will get us out of these tragic situations. The normative landscape is arranged in such a way that we are forced to choose between practical identities with an inescapable regret for not choosing or fulfilling one of our obligations because we are fulfilling another (Williams 1976, 1981, Nussbaum 1985).

With accounts that start from the idea of practical identities and a notion of social reasons, it becomes more difficult to think about reasons that can be straightforwardly weighted against each other. This is because with the multiplicity of social standpoints

<sup>\*</sup> Note that there can still be talk about plurality of values, incommensurable (Raz) and things being 'roughly equal' (Parfit 1987, p. 431). However, that is not what is being claimed here by the talk of tragic. When Raz discusses the plurality of value, he is talking about choices once reasons "have run their course" (Raz 2000, p. 48), for example.

which all generate different obligations, when these obligations call for different or incompatible actions, we are simply unable to act on them all.

This brings us to a potential worry that although the institutional account that has thus far been developed helps us to understand authority as created through a joint commitment, it has said very little about whether or not this authority deserves to be supported by the community (Baumann 2009, pp. 16-17). This is not only because a joint commitment can be coerced into existence but there is also no necessary restriction on what the content of such a joint commitment can be (Gilbert 2006b, pp. 75-82, 228-229). It is clear that a joint commitment is normative in the sense that there is authority and obligation. However, it is not normative in another sense, “a sense standardly connoted by the qualifier 'moral'” (ibid., p. 81.; cf. Gilbert 1996, pp. 299-300, 353 and Gilbert 2000, p. 114-115). This is a particularly important issue to understand so we will examine it at length in terms of Gilbert’s distinction between legitimacy and justice or moral obligation.

It is this distinction between legitimacy and justice which makes it important to emphasise the non-univocity of obligations. We can be obligated to obey the commands of a legitimate authority without being morally obligated to obey these commands. The claim here is that the obligations generated by legitimate authorities are different from the obligations of justice.

Let us begin to clarify this issue. Aristotle, in Book V Chapter 7 of the *Nicomachean Ethics*, develops a distinction between what he calls ‘natural’ justice and legal, or conventional, justice. The first type of justice, ‘natural’ justice, Aristotle tells us, is “that which everywhere has the same force and does not exist by people's thinking this or that” (Aristotle 1984b, 1134b18-19). We must be careful here not to think that what Aristotle is saying is that natural justice is unchanging, for he says “while with us there is



something that is just even by nature, yet all of it is changeable” (ibid. 1134b29-30; cf. Gadamer 2013, p. 330 ). The second type of justice, legal or conventional justice, Aristotle tells us, is “originally indifferent, but when it has been laid down it is not indifferent” (ibid. 1134b20). This part of Aristotle can be difficult to parse but I think there is an important truth to be drawn from it. When we look at norms of a particular institution, it is important to pay attention to *both* the naturally just and the conventionally, or legally, just norms – the latter I would call ‘legitimacy’. However, Aristotle does not give us much on how to evaluate the difference between these norms, and it is rather obscure as to what the importance of this distinction is for him. He seems to mean something like the following: conventional justice is by definition different everywhere and can be evaluated based on natural justice, which is the same everywhere. Aristotle does not tell us what happens with the conflict between natural and conventional justice nor whether conventional justice still has an impact on our practical reasoning in spite of its conflict with natural justice.

A similar distinction has recently been developed by Margaret Gilbert. Her distinction takes place between directed obligation, which is grounded in joint-commitment, and moral obligation. This seems promising and leaves room for a discussion of what Gilbert calls 'the morality of obedience' (Gilbert 2014, p. 426), i.e. when a directed obligation is or is not overridden by moral considerations or perhaps by other directed obligations.

On a phenomenological level, this analysis is highly appealing and makes sense of Stanley Milgram's famous experiment which asks the question of why people feel the pressure, the obligation, to follow immoral commands:

In order to take a close look at the act of obeying, I set up a simple experiment at Yale University... A person comes to a psychological laboratory and is told to carry out a series of acts that come increasingly into conflict with conscience. The main question

is how far the participant will comply with the experimenter's instructions before refusing to carry out the actions required of him (Milgram 1974, pp. 2-3)

If Gilbert's account is correct, the answer is that their directed obligation to obey the commands is in conflict with their moral obligations. Furthermore, this might illuminate why many contemporary theorists have concentrated exclusively on moral obligations to obey the law to the exclusion of directed obligations: there may be anxieties about many people's seeming inability to make the correct decision when their direct and moral obligations are in conflict.

One might hope that all joint commitments are formed with Aristotle's natural justice in mind, as Gilbert has recently remarked:

[i]n particular cases, there could be background understandings or explicit conditions that rule such commands out [i.e., immoral commands], restricting what the parties are jointly committed to... Possibly there is a refined and convincing philosophical argument to the effect that such a proviso is always implicit when relevant joint commitments are made (Gilbert 2014, p. 424).

This might indeed be the case. However, this type of argument runs the risk of moralising, of seeing obligation only in a univocal moral sense, and would miss or distort important aspects of the human condition as Milgram's experiment shows. Further, it is hard to see how such an *a priori* argument would run without a distorting effect on our human practices. It is a fact that people do make immoral commitments to each other and feel obligated to fulfil them. We may not even be aware of our commitments as being immoral.

Milgram's experiment shows one danger of moralising, viz. the influence of directed obligation on the decisions of real people. As Gilbert points out in her "De-Moralizing Political Obligation", by ignoring the motivational significance of directed obligation, we hamper our understanding of political societies and how to deal with morally bad laws (Gilbert 2014, pp. 406-408).

Yet there is another important aspect to this. It is not obvious that all political decisions are going to be reducible to moral decisions, nor is it obvious that moral obligations have a direct bearing on all aspects of joint commitments. Even if it is the case that the moral obligations do have a bearing on all aspects of an institution commitment, it is not obvious from this fact alone whether one ought to abolish or reform the commitment. I am thinking here of John Rawls's opening to his *A Theory of Justice*:

Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust (Rawls 1999a, p. 3).

What I want to draw attention to are the two disjunctions in the above quote, particularly the second: *rejected or revised* and *reformed or abolished*. This is not an insignificant disjunction for our thinking about joint commitments. I take it that this same thought motivates Rawls's later remark that "legitimacy is a weaker idea than justice..." (Rawls 2005 p. 428). Although Rawls is not explicit about this, the later remark may be instructive in understanding the disjunction of the above passage from *A Theory of Justice*. If an institution is unjust *and* illegitimate, we should abolish it, but if it is legitimate despite being unjust we should reform it. Therefore, it is important to grasp what makes an institution or institution commitment legitimate. By moralising obligation, by seeing obligations as obligations only to a just institution, we run the risk of overlooking this important distinction.

Finally, the third, and perhaps the most important, aspect of the importance of directed obligation is to do with the issue of moral disagreement. There are important political issues which are also moral issues that are highly disputed, and yet – politically speaking – an enforceable decision needs to be made. Again, think of the issue of climate

change and the importance of making a decision between a carbon tax model and a Sandel-type fines system. There is a deep moral disagreement here as well as an important practical and political issue. Climate change is an immediate danger to human and animal life and something must be done about it. Yet we should not allow this moral disagreement to invalidate the directed obligation which a state could impose to follow one of these systems. In situations of this sort, the obligations that are part of an institution commitment will help to stabilise these public issues while moral debate continues in the public sphere. It would be problematic for political institutions, or other institutions for that matter, to need to wait for the morally correct answer before committing to a particular policy.

It is generally accepted by most contemporary perspectives that the moral point of view is the dominant point of view for any critique of institutional and political obligation (Horton 20010, p. 144 and Knowles 2009, p. 188). The idea here is that the obligations which count and make one actually accountable to an institution are those that can be shown to be moral obligations. A major problem with this point of view, from my perspective, is the assumption that underlies it. The assumption seems to be that we have some idea of what a 'final' morality will look like or what approximates it, that we have access to a privileged point of view outside our present context. However, when one surveys contemporary political and moral debates, one will at once notice entrenched disagreements on virtually every moral issue. For example, in debates about political authority and political obligations, there are deep disagreements between philosophical anarchists, natural duties theorists (fair-play, gratitude and Samaritanism) and associative obligations theorists. There is even the emergence of pluralistic accounts which combine these different views in interesting and compelling ways. However, the point I would like to make is that there does not seem to be any forthcoming agreement on this issue which one

needs if there is going to be a knock-down argument from the moral point of view. This is not to suggest that these views do not give us resources and important insights to critique current political authorities. They do. Rather, the point is that neglecting other important aspects of our practical lives may be detrimental to further progress on these issues.

I think that there are three general perspectives in which a critique of the content of a command can be given. The first is by far the most dominant in contemporary political and moral philosophy, e.g. universally valid norms or what Aristotle has called ‘natural’ justice. Analogous to this perspective is that it is from a privileged epistemic point of view, e.g. climate scientists’ privileged view point on global climate change. I will not concentrate on this perspective, but everything said about the moral point of view will, *mutatis mutandis*, also hold for the epistemic, unless otherwise noted. These two points of view can be considered forms of transcendent critique because they are points of view outside specific commitments of a particular institution. When the content of these commitments is absolutely unjustified in terms of one of these points of view, the directed obligation to conform is overridden. That is, the institutional commitment still exerts force on its members but, from the moral point of view, the obligation to the institution should not be followed.

Why should we think that directed obligations are overridable and not, as Stephen Darwall thinks, as “defeasible – so defeasible, in fact, that it may never have any force at all...” (Darwall 2006, p. 186*n*9)? One way to make sense of this dispute is that when a reason is overridden, there is some remainder left over, some normative reason to feel remorse. There may be reasons for one to justify oneself to the group for breaking the directed obligation. When a reason is defeasible, there is no remainder. The reason is defeated. There are two types of response we can make to Darwall.

The first response is that it is not entirely clear why the fact that a reason has a very weak force means that there will be no remainder or why a justification cannot be demanded by the institution's members for the non-conformity. Furthermore, it is not entirely clear why we are to think that the forces of directed obligation are so weak. It seems to be a bare assertion that is being made with no forthcoming justification.

The second response is epistemic. We cannot rule out the possibility of error. In our current epistemic position, when the commitments of an institution come into conflict with what we take to be morally right, it may bring about the process of moral learning. It may lead us to revise what we take to be morally right. This is part of the advantage of taking seriously the potential conflicts between directed obligations and moral obligations. In this way, the overrideability, as opposed to defeasibility, may keep us from disregarding our directed obligations too hastily.

To understand why it is important to understand the difference between defeasibility and overrideability, let us look at three different types of case. The way we can make sense of this is that having reasons not to conform does not undermine the genuineness of the obligation – genuine obligations are not *pro tanto* (so far as it goes) obligations. In Gilbert's words:

if an obligation of this type is present in one context (e.g., an agreement stands), then it does not disappear if the context is enlarged (e.g., the agreement still stands, but one can save someone's life by violating it). It may be discounted in the light of the additional considerations, but it does not disappear (Gilbert 1996, p. 299).

Although each member of the institution is obligated to the institution qua member, each member also transcends the institution qua individual (or member of another institution). It is from the transcendent standpoint, an enlarged context, that the group must be justified in its demands. This transcendent standpoint does not, however, need to be a 'view from

nowhere'. We can identify at least three standpoints (Habermas 1998, Ch. 4.2, Habermas 1994, Ch. 1, Forst 2012, Ch. 3 and Forst 2002, Ch. 5).

First, we have what we can term 'instrumental' cases. In these cases, an institution may be unjustified if, for example, in the case of collective belief (Gilbert 1992 Ch. V, Gilbert 1996 Ch. 14), the belief is false or, in the case of a goal, the institution's plan is not the most effective course of action. In these types of case, members may not necessarily be epistemic equals. Say, on our walk together where you have the standing to decide in which direction to go, we come upon a fork in the road. You say "Left!" but I know the left is a dead-end and we want to have a long walk. In this case, although I do not have the proper standing, I still would be justified in saying "No, right! Left is a dead-end." In these 'instrumental' cases, the knowledge of the individuals transcends that of the institution and may be justified in correcting or, in extreme cases, even dissolving a plural subject.

The fact that I may be justified in doing so does not automatically make the institution or the obligations which I have to it void. Rather, it puts me in the position to try to convince you that you are going the wrong way. This may indeed cause conflict and tension within the group because you could still rightfully claim that you are the one leading the group and that I am obligated by that fact. This would be true, but the instrumental considerations may override the obligations I have to the group and you should, for this wider perspective, change the direction you are taking us.

Next is from the individual point of view, e.g. the suffering of the individual who is deprived of basic capabilities. Amartya Sen gives the following examples of basic capabilities: "the ability to be well-nourished and well-sheltered, the capability of escaping avoidable morbidity and premature mortality and so forth" (Sen 1992, pp. 44-45). It is no surprise that these basic capabilities are part of our moral commitments. The case I have in

mind here is when the moral point of view is in agreement with certain institutional commitments but still leads to the suffering of the individual. It is only the “cries of the wounded” (James 1979, p. 158), to borrow a phrase from William James, which can bring our attention to individual suffering.

We have cases of 'moral obligation'. In this type of case, there is some conflict between an obligation incurred in virtue of joint commitments and some “intrinsic good” or “moral norm” (for our present purpose we can remain agnostic about the nature of “intrinsic goods” or “moral norms”). Take, for example, an (in)famous case that occurs in Hobbes:

The question is often asked whether *agreements* extorted by fear are obligatory or not. For example, am I obligated if, to save my life, I make an *agreement* with a highway robber to pay him thousand gold pieces tomorrow, and to do nothing that might result in his arrest and arraignment (Hobbes 1998, p. 38)

Now, we can agree with Hobbes that from the perspective of the institution that is formed by this agreement one is obligated to pay the highway robber, but from the perspective of 'morality' we are justified in taking action to have him arrested – that is, to *break* our obligation. As Hobbes notes, this is only the case in the state of nature (ibid.). Let us then take a non-state of nature example.

Some paradigmatic cases for this could be: a pharmacist who opposes abortion on moral grounds in a state which has a law that obligates her to sell abortifacient drugs to patients with prescriptions, refusal to follow the Jim Crow laws of the southern United States after Reconstruction (1877-1965) or refusing to obey orders to perform “enhanced interrogation techniques”. These can all be classed as refusal based on moral obligations. Whether they are all cases of “true” or justified objections we can leave to the side. The important point is that there is a place to be a moral objector to state obligation.



I think it would be an exaggeration if I were to say that there are only conflicts between moral and individual perspectives in regard to institutions and not between conflicting institutional perspectives themselves. We have what we can call cases of individual 'bifurcated obligation'. This is due to the intuitive idea that individuals can and do belong to more than one institution at a time (Gilbert 1992, p. 220; cf. Forst 2002 p. 272). These situations are quite familiar due to their sometimes tragic nature, e.g. Sophocles's *Antigone* (cf. Hollis 1996, Ch. 7 and 9). However, there can be less dramatic examples: Jack and Jill are parties of a joint commitment to have dinner together. As soon as they arrive at the restaurant, Jill receives a call from her boss informing her that there is an emergency which she needs to take care of immediately (we will stipulate that this is part of her normal obligations at her place of work, though it rarely occurs). In this situation, it is clear that Jill is obligated both to stay to eat dinner with Jack *and* leave to take care of the work emergency. Jack as well as Jill's boss both have the standing to rebuke her from different points of view. Neither of her commitments are predicated on a conditional commitment. She has not jointly committed to have dinner together with Jack if, and only if, she is not called into work – she has an unconditional commitment to have dinner with Jack until such time as the commitment is fulfilled or dissolved. Now as it happens, Jack is has a sympathetic nature and realises the consequences of Jill not immediately leaving. Jack and Jill jointly agree to postpone their dinner together – though Jill may still feel remorse or regret. The important point here is that obligations can conflict and, as Bernard Williams has pointed out, they “are neither systematically avoidable, nor all soluble without remainder” (Williams 1973. p. 179).

Similar cases exist in the political sphere. Think of a person who grew up as a Quaker (Rawls 2005, p. 393), a group known for being conscientious objectors, but who is

not particularly religious but remains deeply committed to her/his family and is also deeply patriotic to their country. During war-time this person is drafted, hence has the obligation to serve. The family may, on the other hand, rebuke him or her for not being a conscientious objector. There seems to be no non-obligation breaking decision for this person to make.

This is the second type of case. We can see this in terms of a de-centring of the subject. Famously, W.E.B Du Bois described this type of de-centring in the American experience in 1903 as the 'peculiar sensation' of 'double-consciousness':

this sense of always looking at one's self through the eyes of others, of measuring one's soul by the tape of a world that looks on in amused contempt and pity. One ever feels his twoness – an American, a Negro; two souls, two thoughts, two unreconciled strivings; two warring ideals in one dark body, whose dogged strength alone keeps it from being torn asunder. The history of the American Negro is the history of this strife,—this longing to attain self-conscious manhood, to merge his double self into a better and truer self. In this merging he wishes neither of the older selves to be lost. He would not Africanize America, for America has too much to teach the world and Africa. He would not bleach his Negro soul in a flood of white Americanism, for he knows that Negro blood has a message for the world. He simply wishes to make it possible for a man to be both a Negro and an American, without being cursed and spit upon by his fellows, without having the doors of Opportunity closed roughly in his face (Bois 1994, pp. 2-3; cf. Anderson 2001, pp. 35-36).

Although there are clear political tones to Du Bois's words, to see them as merely political would be a simplification and a falsification. Du Bois's point is much further-reaching than the political. The idea is that because we have a multiplicity of practical identities due to our membership of different institutions, we can play them off one another in order to improve all of the institutions.

Since we feel committed to furthering and conforming to what we find desirable and compelling about our different identities, we feel not only the motivational force but also the normative force of conforming to them. However, when one identity conflicts with another we can be motivated to attempt to make both of these identities and the institutions from which they are derived compatible with each other.

There is the internal point of view which makes up the third type of case and is different from the above because it is a case of defensibility, rather overridability. As opposed to the above positions, when a command is in conflict with a constitutive commitment, there is no directed obligation at all. One does not owe the institution conformity to commitments which violate its constitutive rules. From the point of view of the institution, our fundamental obligation is conformity to the constitutive commitments. It would be natural to say that it is outside the standing of the authorising members to command such things. This might confuse matters. The issue here is not one of standing but one of content. The United States Congress, presumably, has the standing to make laws in the United States. However, it does not have the authority to issue commands that violate the US Constitution. This is, strictly speaking, an issue of content and not of standing as was argued above.

Contra Simmons, who thinks an account based on Gilbert-style joint-commitments encourages confused, oppressed, unthinking people to make “leaps of faith” (Simmons 2001, p. 75), Gilbert writes:

I in no way urge unreflective commitment. On the contrary, that is all too easy to come by, and should be challenged, and become – if we were being reasonable all along – reflective commitment (Gilbert 1996, p. 372)

We can read Gilbert's comments about reflective commitment as a call to redeem our 'genuine' obligations in regard to the above 'contexts of justification'. As is clear, each of the above contexts (instrumental/pragmatic, ethical and moral) have their own justification grammars (Habermas 1996 and Forst 2002). Furthermore, it seems clear that for the most part, the call for redemption of obligation is prompted by conflict and disagreement, as demonstrated by the cases above (Honneth 1995). That is, the conflict between “our wants and my wants” (Gilbert 1992 pp. 424-425) – although self and

social-reflection as per Rawls's original position thought experiment and reflective equilibrium are other important avenues (Rawls 2001, §§6 and 10). However, we should not overlook the fact that Rawls also sees the positive side of conflict. His analysis in *Political Liberalism* of the origins of liberal tolerance in the religious wars of tolerance, which he credits Hegel with also seeing, is a paradigm example of the positive aspect of social conflict.

Let us take a look at conflict between obligations and how this can lead to progress and moral learning. We are all familiar with Antigone's dilemma: she must decide whether to follow her familial duty and bury her traitor brother Polynices, or follow King Creon's edict forbidding the burial of a traitor, i.e. her political duty. Today, we have a strong tendency to reconcile these two conflicting demands through formalistic theories. That is, we tend to think the only way to view this sort of issue is from the moral point of view, that the only reasons that matter are moral reasons (Williams 1981 Ch. 1 and Scheffler 1994).

Charles Taylor put it best when he said:

that the price of modern formalism [...] has been a severe distortion of our understanding of our moral thinking. One of the big illusions which grows from [this reduction] is the belief that there is a single consistent domain of the 'moral', that there is one set of considerations, or mode of calculation, which determines what we ought 'morally' to do (Taylor 1985, p. 233).

The 'modern formalism' which Taylor is identifying is a type of reductionism which discounts other points of view, e.g. joint commitments.

Instead of this modern formalism, we must acknowledge Rainer Forst's point that:

[a] form of life does not become false because its general realization cannot be morally demanded. Just as ethical reasons are not necessarily general reasons in a moral sense, moral reasons are not sufficient to determine the good life (Forst 2002, p. 39).

This is precisely the point that is being argued here. We need to pay attention to joint

commitments. It certainly is the case that the most joint commitments are not universalisable.

What we end up with is a situation like the following. X qua institution member has sufficient reason to  $\phi$  in situation S, all things being equal for X qua institution member. However, X qua moral being does not have sufficient reason to  $\phi$  in situation S, all things being equal (Anderson 2001). This is Simmel's point about individuals transcending their society. This does not make the sufficient reason superfluous internal to an institution. For one thing, it helps to solve coordination problems – e.g. it is what makes driving on the correct side of the road rational (Gilbert 2006b, p. 33).

In the above mentioned political cases, we need a theory which helps us determine when the state is justified in using coercive power to enforce its policies against 'instrumental' claims, 'value/ethical' claims and 'moral' claims. However, there is a problem with a sub-class of 'instrumental' cases with which deliberative democracy in its standard form cannot deal. This has to do with what Habermas calls “administrative power” which the deliberating public cannot itself rule but only point “administrative power in specific directions” (Habermas 1996, p. 300).

Gilbert has given us an understanding of the proper standing necessary for someone *in* authority to be able to create sufficient reasons, although perhaps in a limited fashion, for those commanded. X has authority over society S if, and only if, there is a joint commitment in society S to accept the commands of X where a joint commitment is characterised by: (1) an expression of readiness to accept Y together on the part of all in society S and (2) it is common knowledge in society S that 1 exists.

This can be seen clearly in the rise of religious tolerances to a moral norm through the violence of the wars of religion. According to Rawls, liberalism and its value of

tolerance, came directly out of the Thirty Years War which was concerned with the power of the Holy Roman Empire (Rawls 2005, p. xxvi). The *modus vivendi* that emerged from the religious wars has now become solidified into moral principles: the interconnected principles of freedom of conscience and tolerance. As Rawls notes, Hegel might have been the first to acknowledge this point. In his *Philosophy of Right*, Hegel said:

This division [in the church] is likewise the most fortunate thing which could have happened to the church and to thought as far as their freedom and rationality are concerned (Hegel 2008, §270 Remark).

He quite clearly sees that religious pluralism was a precondition for religious liberty. It is in this way that we can see the truth in Raymond Geuss's recent claim that:

Ethics is usually dead politics: the hand of a victor in some past conflict reaching out to try to extend its grip to the present and the future. There is nothing inherently wrong with this. Our past is an essential part of what we are, which we ignore at our peril (Geuss 2009, p. 42).

What we can bring out of these considerations is that we need to integrate all three perspectives into our evaluative process. We need to critique our institutions simultaneously from the transcendent point of view and the immanent point of view while paying close attention to the implications of both critiques.

We can say, in the tradition of Nelson Goodman and John Rawls, a moral obligation is amended if it conflicts with a directed obligation we are unwilling to amend; a directed obligation is amended if it violates a moral obligation we are unwilling to amend. The process of justification is a delicate one of making mutual adjustments between moral obligations and directed obligations, and the needs of the individual. The agreement achieved lays the only justification needed for either (Goodman 1983, §2 and Rawls 1999a, p. 18n7). This, then, is a three dimensional view of justification. That is, there will be a process of reaching reflective equilibrium between: (1) groups and the moral, (2) groups

and the individual, (3) the moral and the individual.

One way to see what is going on here is the 'critical role' that is constitutive of the creation of an agent's practical identity (Korsgaard 1996b p. 101) – to make oneself intelligible as an agent. This is our second-nature which separates merely natural beings from culturally bathed humans. There are all sorts of roles which we are more or less thrown into: subject qua moral community, subject qua political community, subject qua family member, subject qua employee, etc. Part of constructing oneself as an agent, then, is, in the words of Korsgaard:

The work of pulling ourselves back together is also the work of pulling those identities into a single practical identity, choosing among them when we have to, deciding which is to have priority, harmonizing them when we can (Korsgaard 2009, p. 126; cf Brandom 2009, p. 52)

The recognition of the unreconciled state of individuals, i.e. conflicts that arise between different roles to which one is committed, is an impetus to the critical stance (cf. Adorno 2005, p. 39 and Horkheimer 2004, pp. 76-77). As Simone Weil once put it:

No human being, whoever he may be, under whatever circumstances, can escape them [obligations] without being guilty of crime; save where there are two genuine obligations which are in fact incompatible, and a man is forced to sacrifice one of them.

The imperfections of a social order can be measured by the number of situations of this kind it harbours within itself (Weil 2002, p. 3).

## Bibliography

- Adorno, Theodor W. 1973. *Negative Dialectics*. Translated by E.B. Ashton. London: Routledge.
- . 2001. *Problems of Moral Philosophy*. Edited by Thomas Schröder. Translated by Rodney Livingstone. Stanford: Stanford University Press.
- . 2005. *Minima Moralia: Reflections on a Damaged Life*. Translated by E.F.N. Jephcott. London: Verso.
- . 2006. *History and Freedom: Lectures 1964-1965*. Edited by Rolf Tiedemann. Translated by Rodney Livingstone. Cambridge: Polity.
- Agamben, Giorgio. 2008. *State of Exception*. Translated by Kevin Attell. Chicago: University of Chicago Press.
- Anderson, Benedict. 2006. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised Edition. London: Verso.
- Anderson, Elizabeth. 2000. "Beyond Homo Economicus: New Developments in Theories of Social Norms." *Philosophy & Public Affairs* 29 (2): 170–200.
- . 2001. "Symposium on Amartya Sen's Philosophy: 2 Unstrapping the Straitjacket of : A Comment on Amartya Sen's Contributions to Philosophy and Economics." *Economics and Philosophy* null (01): 21–38.
- Anscombe, G.E.M. 2000. *Intention*. 2nd ed. Cambridge, MA: Harvard University Press.
- Apel, Karl-Otto. 1996. *Selected Essays: Ethics and the Theory of Rationality*. Edited by Eduardo Mendieta. Translated by Glyn Adey and David Frisby. Vol. 2. New Jersey: Humanities Press.
- Arendt, Hannah. 2000. "What Is Authority?" In *The Portable Hannah Arendt*, edited by Peter Baehr. New York: Penguin Books.
- Aristotle. 1984a. *Metaphysics*. In *The Complete Works of Aristotle*, edited by Jonathan Barnes, translated by W.D. Ross, The Revised Oxford Translation. Vol. 2. Princeton: Princeton University Press.
- . 1984b. *Nicomachean Ethics*. In *The Complete Works of Aristotle*, edited by Jonathan Barnes, translated by W.D. Ross and J.O. Urmson, The Revised Oxford Translation. Vol. 2. Princeton: Princeton University Press.
- . 1984c. *Politics*. In *The Complete Works of Aristotle*, edited by Jonathan Barnes, translated by B. Jowett, The Revised Oxford Translation. Vol. 2. Princeton: Princeton University Press.
- Bauman, Zygmunt. 1989. *Modernity and the Holocaust*. Cambridge: Polity.
- Baumann, Caroline M. 2009. "Reconsidering Gilbert's Account of Social Norms." In *EPSA Epistemology and Methodology of Science*, edited by Mauricio Suárez, Mauro Dorato, and Miklós Rédei, 13–22. Springer Netherlands.
- Benhabib, Seyla. 1994. "Deliberative Rationality and Models of Democratic Legitimacy." *Constellations* 1 (1): 26–52.
- Bloom, Harold. 2001. *How to Read and Why*. New York: Simon and Schuster.
- Bohman, James. 2000. *Public Deliberation: Pluralism, Complexity, and Democracy*. Cambridge, MA: MIT Press.
- Bois, W. E. B. Du. 1994. *The Souls of Black Folk*. Mineola: Dove.
- Brandom, Robert. 1979. "Freedom and Constraint by Norms." *American Philosophical Quarterly* 16 (3): 187–96.



- . 1998. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- . 2001. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- . 2007. “The Structure of Desire and Recognition: Self-Consciousness and Self-Constitution.” *Philosophy and Social Criticism* 33 (1): 127–50.
- . 2009. *Reason in Philosophy: Animating Ideas*. Cambridge, MA: Harvard University Press.
- . 2013. “Some Post-Davidsonian Elements of Hegel’s Theory of Agency.” In *Sinnkritisches Philosophieren*, edited by Sebastian Rödl and Henning Tegtmeier. Berlin: De Gruyter.
- Bratman, Michael. 1999a. *Intention, Plans, and Practical Reason*. Stanford: CSLI Publications.
- . 1999b. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.
- . 2007. *Structures of Agency: Essays*. Oxford: Oxford University Press.
- . 2012. “Time, Rationality, and Self-Governance.” *Philosophical Issues* 22 (1): 73–88.
- . 2014. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.
- Broome, John. 2007. “Wide or Narrow Scope?” *Mind* 116 (462): 359–70.
- Buber, Martin. 1996. *I And Thou*. Translated by Walter Kaufmann. New York: Simon & Schuster.
- . 2002a. “Afterword: The History of the Dialogical Principle.” In *Between Man and Man*, translated by Maurice Friedman. London: Routledge.
- . 2002b. “What Is Man?” In *Between Man and Man*, translated by Ronald Gregor-Smith. London: Routledge.
- Caney, Simon, and Cameron Hepburn. 2011. “Carbon Trading: Unethical, Unjust and Ineffective?” *Royal Institute of Philosophy Supplements* 69 (October): 201–34.
- Carman, Taylor. 2003. *Heidegger’s Analytic: Interpretation, Discourse and Authenticity in Being and Time*. Cambridge University Press.
- Cavell, Stanley. 1999. *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy*. New Edition. Oxford: Oxford University Press.
- Cicero, Marcus Tullius. 2000. “Pro Roscio Amerino.” In *Defence Speeches*, translated by D.H. Berry. Oxford: Oxford University Press.
- Cohen, Joshua. 1997. “Deliberation and Democratic Legitimacy.” In *Deliberative Democracy: Essays on Reason and Politics*, edited by James Bohman and William Rehg. Cambridge, MA: MIT Press.
- Dancy, Jonathan. 2000. *Practical Reality*. Oxford: Oxford University Press.
- Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
- . 2010. “Authority and Reasons: Exclusionary and Second-Personal.” *Ethics* 120 (2): 257–78.
- . 2013a. *Morality, Authority, and Law: Essays in Second-Personal Ethics I*. Oxford: Oxford University Press.
- . 2013b. *Honor, History, and Relationship: Essays in Second-Personal Ethics II*. Oxford: Oxford University Press.

- Davidson, Donald. 2001. "Actions, Reasons, and Causes." In *Essays on Actions and Events*, Second Edition. Oxford: Oxford University Press.
- Derrida, Jacques. 1989. "Force of Law: The 'Mystical Foundation of Authority.'" Translated by Mary Quantance. *Cardozo Law Review* 11: 920.
- Dunn, John. 1996. "Political Obligation." In *The History of Political Theory and Other Essays*. Cambridge: Cambridge University Press.
- . 2000. *The Cunning of Unreason: Making Sense of Politics*. London: HarperCollins.
- Dewey, John. 1983. *Human Nature and Conduct*. Edited by Jo Ann Boydston. Carbondale: Southern Illinois University Press.
- . 2008. *Logic: The Theory of Inquiry*. Edited by Jo Ann Boydston. Carbondale: Southern Illinois University Press.
- Durkheim, Émile. 2013. *The Rules of Sociological Method: And Selected Texts on Sociology and Its Method*. Edited by Steven Lukes and W.D. Halls. Third Edition. New York: Simon and Schuster.
- Dworkin, Ronald. 1986. *Law's Empire*. Cambridge, MA: Harvard University Press.
- Edmundson, William A. 2013. "Because I Said So." *Problema: Anuario de Filosofía Y Teoría Del Derecho* 7: 41–61.
- Engstrom, Stephen. 2009. *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge, MA: Harvard University Press.
- Enoch, David. 2009. "On Estlund's 'Democratic Authority.'" *Iyyun: The Jerusalem Philosophical Quarterly* 58 (January): 35–48.
- . 2013. "The Disorder of Public Reason." *Ethics* 124 (1): 141–76.
- . 2014. "Authority and Reason-Giving." *Philosophy and Phenomenological Research* 89 (2): 296–332.
- Estlund, David. 1997. "Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority." In *Deliberative Democracy: Essays on Reason and Politics*, edited by James Bohman and William Rehg. Cambridge, MA: MIT Press.
- . 2008. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press.
- Feuerbach, Ludwig. 2012. *Principles of the Philosophy of the Future*. In *The Fiery Brook: Selected Writings*, edited and translated by Zawar Hanfi. London: Verso Books.
- Forst, Rainer. 2002. *Contexts of Justice: Political Philosophy Beyond Liberalism and Communitarianism*. Translated by John M. M. Farrell. Berkeley: University of California Press.
- . 2012. *The Right to Justification: Elements of a Constructivist Theory of Justice*. Translated by Jeffrey Flynn. New York: Columbia University Press.
- . 2014. *Justification and Critique: Towards a Critical Theory of Politics*. Translated by Ciaran Cronin. Cambridge: Polity Press.
- Frankfurt, Harry G. 1998. *The Importance of What We Care About: Philosophical Essays*. Cambridge, MA: Cambridge University Press.
- Freeman, Samuel. 1990a. "Constitutional Democracy and the Legitimacy of Judicial Review." *Law and Philosophy* 9 (4): 327–70.
- . 1990b. "Reason and Agreement in Social Contract Views." *Philosophy & Public Affairs* 19 (2): 122–57.
- Friedman, Richard B. 1989. "Oakeshott on the Authority of Law." *Ratio Juris* 2 (1): 27–40.
- . 1990. "On the Concept of Authority in Political Philosophy." In *Authority*, edited

- by Joseph Raz. New York: New York University Press.
- Gadamer, Hans-Georg. 2013. *Truth and Method*. Translated by Joel Weinsheimer and Donald G. Marshall. Second, Revised Edition. London: Bloomsbury.
- Gaus, Gerald. 1997. "Reason, Justification, and Consensus: Why Democracy Can't Have It All." In *Deliberative Democracy: Essays on Reason and Politics*, edited by James Bohman and William Rehg. Cambridge, MA: MIT Press.
- . 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.
- Gauthier, David P. 1986. *Morals by Agreement*. Oxford: Oxford University Press.
- Geuss, Raymond. 2001. *History and Illusion in Politics*. Cambridge: Cambridge University Press.
- . 2008. *Philosophy and Real Politics*. Princeton: Princeton University Press.
- . 2009. *Outside Ethics*. Princeton: Princeton University Press.
- Gilbert, Margaret. 1992. *On Social Facts*. Princeton: Princeton University Press.
- . 1996. *Living Together: Rationality, Sociality, and Obligation*. Plymouth: Rowman & Littlefield.
- . 2000. *Sociality and Responsibility: New Essays in Plural Subject Theory*. Lanham: Rowman & Littlefield.
- . 2003. "The Structure of the Social Atom: Joint Commitment as the Foundation of Human Social Behavior." In *Socializing Metaphysics: The Nature of Social Reality*, edited by Frederick Schmitt. Lanham: Rowman & Littlefield Publishers.
- . 2006a. "Who's to Blame? Collective Moral Responsibility and Its Implications for Group Members." *Midwest Studies in Philosophy* 30 (1): 94–114.
- . 2006b. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: Oxford University Press.
- . 2014. *Joint Commitment: How We Make the Social World*. Oxford: Oxford University Press.
- Goodman, Nelson. 1983. *Fact, Fiction, and Forecast*. Fourth Edition. Cambridge, MA: Harvard University Press.
- Green, Leslie. 1988. *Authority of the State*. Oxford: Clarendon Press.
- Greenwood, John D. 2003. "Social Facts, Social Groups and Social Explanation." *Noûs* 37 (1): 93–112.
- Gutmann, Amy, and Dennis Thompson. 2004. *Why Deliberative Democracy?*. Princeton: Princeton University Press.
- Habermas, Jürgen. 1984. *The Theory of Communicative Action: Reason and the Rationalization of Society*. Translated by Thomas McCarthy. Vol. 1. 2 vols. Boston: Beacon Press.
- . 1992a. *Moral Consciousness and Communicative Action*. Translated by Christian Lenhardt and Shierry Weber Nicholsen. Cambridge: Polity Press.
- . 1992b. *Postmetaphysical Thinking: Philosophical Essays*. Translated by William Mark Hohengarten. Cambridge: Polity Press.
- . 1995. *Justification and Application: Remarks on Discourse Ethics*. Translated by Ciaran Cronin. Cambridge: Polity Press.
- . 1998. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Translated by William Rehg. Cambridge: MIT Press.
- . 1999. *The Inclusion of the Other: Studies in Political Theory*. Edited by Ciaran Cronin and Pablo De Greiff. Cambridge: Polity Press.

- Hampshire, Stuart. 1989. *Innocence and Experience*. London: Penguin Books.
- Hampton, Jean. 1986. *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
- . 1997. *Political Philosophy*. Boulder: Westview Press.
- Hart, H. L. A. 1955. “Are There Any Natural Rights?” *The Philosophical Review* 64 (2): 175–91.
- . 2012. *The Concept of Law*. Third Edition. Oxford: Oxford University Press.
- Hegel, G.W.F. 1977. *Phenomenology of Spirit*. Translated by A.V. Miller. Oxford: Oxford University Press.
- . 2008. *Outlines of the Philosophy of Right*. Edited by Stephen Houlgate. Translated by T.M. Knox. Oxford: Oxford University Press.
- Heidegger, Martin. 1962. *Being and Time*. Translated by John Macquarrie and Edward Robinson. New York: Harper & Row.
- Held, David. 2006. *Models of Democracy*. Third Edition. Cambridge: Polity Press.
- Hershovitz, Scott. 2003. “Legitimacy, Democracy and Razian Authority.” *Legal Theory* 9 (03): 201–20.
- . 2011. “The Role of Authority.” *Philosophers’ Imprint* 11 (7).
- Hersh, Seymour M. 2015. “The Scene Of The Crime: A Reporter’s Journey to My Lai and the Secrets of the Past.” *New Yorker*, March 30.
- Hindriks, Frank. 2003. “Social Groups, Collective Intentionality, and Anti-Hegelian Skepticism.” In *Realism in Action*, edited by Matti Sintonen, Petri Ylikoski, and Kaarlo Miller, 213–31. Synthese Library 321. Springer Netherlands.
- Hobbes, Thomas. 1996. *Leviathan*. Edited by Richard Tuck. Revised Student Edition. Cambridge, MA: Cambridge University Press.
- . 1998. *On the Citizen*. Edited by Richard Tuck and Michael Silverthorne. Cambridge: Cambridge University Press.
- Hollis, Martin. 1996. *Reason in Action: Essays in the Philosophy of Social Science*. Cambridge: Cambridge University Press.
- Honneth, Axel. 1996. *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Translated by Axel Anderson. Cambridge: Polity Press.
- . 2012. *The I in We: Studies in the Theory of Recognition*. Cambridge: Polity Press.
- Honneth, Axel, and Hans Joas. 1988. *Social Action and Human Nature*. Translated by Raymond Meyer. Cambridge: Cambridge University Press.
- Horkheimer, Max. 2004. *Eclipse of Reason*. London: Continuum.
- Horton, John. 2006. “In Defence of Associative Political Obligations: Part One.” *Political Studies* 54 (3): 427–43.
- . 2007. “In Defence of Associative Political Obligations: Part Two.” *Political Studies* 55 (1): 1–19.
- . 2010. *Political Obligation*. Second Edition. London: Palgrave Macmillan.
- Houlgate, Stephen. 2013. *Hegel’s “Phenomenology of Spirit”: A Reader’s Guide*. London: Bloomsbury.
- Hume, David. 1975. *An Enquiry Concerning the Principles of Morals*. In *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by P. H. Nidditch, 3rd Edition. Oxford: Clarendon Press.
- . 1985. “Of the Original Contract.” In *Essays, Moral, Political, and Literary*, edited by Eugene F. Miller, Revised Edition. Indianapolis: Liberty Classics.
- James, Aaron. 2007. “Legal and Other Governance in Second-Person Perspective.” *Loyola*

- of *Los Angeles Law Review* 40 (3): 911.
- James, William. 1979. "The Moral Philosopher and Moral Life." In *The Will to Believe, and Other Essays in Popular Philosophy*, edited by Frederick Burkhardt, Fredson Bowers, and Ignas K. Skrupskelis. The Works of William James. Cambridge, MA: Harvard University Press.
- Kant, Immanuel. 1996a. *Critique of Practical Reason*. In *Practical Philosophy*, edited and translated by Mary J. Gregor. Cambridge: Cambridge University Press.
- . 1996b. *Groundwork of the Metaphysics of Morals*. In *Practical Philosophy*, edited and translated by Mary J. Gregor. Cambridge: Cambridge University Press.
- . 1996c. *The Metaphysics of Morals*. In *Practical Philosophy*, edited and translated by Mary J. Gregor. Cambridge: Cambridge University Press.
- . 1998. *Critique of Pure Reason*. Edited and translated by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.
- . 2004. *The Jäsche Logic*. In *Lectures on Logic*, edited and translated by J. Michael Young. Cambridge: Cambridge University Press.
- . 2007. *Anthropology from a Pragmatic Point of View*. In *Anthropology, History, and Education*, edited by Günter Zöllner and Robert B. Louden. Cambridge: Cambridge University Press.
- Klosko, George. 1987. "The Principle of Fairness and Political Obligation." *Ethics* 97 (2): 353–62.
- Knowles, Dudley. 2009. *Political Obligation: A Critical Introduction*. London: Routledge.
- Kompridis, Nikolas. 2011. *Critique and Disclosure: Critical Theory Between Past and Future*. Cambridge, MA: MIT Press.
- Korsgaard, Christine M. 1996a. *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge, MA: Cambridge University Press.
- . 1996b. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- . 2007. "Autonomy and the Second Person Within: A Commentary on Stephen Darwall's the Second-Person Standpoint." *Ethics* 118 (1): 8–23.
- . 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- . 2009a. "The Activity of Reason." *Proceedings and Addresses of the American Philosophical Association* 83 (2): 23–43.
- . 2009b. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- . 2014. "The Normative Constitution of Agency." In *Rational and Social Agency: The Philosophy of Michael Bratman*, edited by Manuel Vargas and Gideon Yaffe. Oxford: Oxford University Press.
- Laden, Anthony Simon. 2000. "Outline Of a Theory Of Reasonable Deliberation." *Canadian Journal of Philosophy* 30 (4): 551–79.
- . 2005. "Evaluating Social Reasons: Hobbes versus Hegel." *The Journal of Philosophy* 102 (7): 327–56.
- Lavin, Douglas. 2014. "Other Wills: The Second-Person in Ethics." *Philosophical Explorations*, no. 3: 1–10.
- Lear, Jonathan. 1982. "Leaving the World Alone." *The Journal of Philosophy* 79 (7): 382–403.
- . 1984. "The Disappearing 'We.'" *Aristotelian Society Supplementary Volume* 58: 219–42.



- Lewis, David. 2002. *Convention: A Philosophical Study*. Oxford: Blackwell Publishers.
- List, Christian, and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.
- Locke, John. 1993. *The Second Treaties of Government*. In *Locke: Political Writings*, edited by David Wootton. Indianapolis: Hackett Publishing.
- MacIntyre, Alasdair. 2011. *After Virtue: A Study in Moral Theory*. Third Edition. London: Bristol Classical Press.
- Manin, Bernard. 1987. "On Legitimacy and Political Deliberation." Translated by Elly Stein and Jane Mansbridge. *Political Theory* 15 (3): 338–68.
- Marmor, Andrei. 2009. *Social Conventions: From Language to Law*. Princeton: Princeton University Press.
- . 2011a. "An Institutional Conception of Authority." *Philosophy & Public Affairs* 39 (3): 238–61.
- . 2011b. "The Dilemma of Authority." *Jurisprudence* 2 (1): 121–41.
- Marx, Karl. 2000. "Grundrisse." In *Karl Marx: Selected Writings*, edited and translated by David McLellan, Second Edition. Oxford: Oxford University Press.
- McDowell, John. 1996. *Mind and World*. Cambridge, MA: Harvard University Press.
- . 1998. "Two Sorts of Naturalism." In *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
- Mead, George Herbert. 1962. *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. Edited by Charles W. Morris. Chicago: University of Chicago Press.
- Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. London: Tavistock Publications.
- Miller, Gregg Daniel. 2011. *Mimesis and Reason: Habermas's Political Philosophy*. SUNY Press.
- Montaigne, Michel. 2004. "On Experience." In *The Complete Essays*, edited and translated by M.A. Screech. London: Penguin Books Limited.
- Nussbaum, Martha. 1985. "Aeschylus and Practical Conflict." *Ethics* 95 (2): 233–67.
- Oakeshott, Michael. 1990. *On Human Conduct*. Oxford: Clarendon Press.
- O'Neill, Onora. 2000. *Bounds of Justice*. Cambridge: Cambridge University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- . 2013. *On What Matters*. 2 vols. Oxford: Oxford University Press.
- Peter, Fabienne. 2007. "Democratic Legitimacy and Proceduralist Social Epistemology." *Politics, Philosophy & Economics* 6 (3): 329–53.
- . 2008. "Pure Epistemic Proceduralism." *Episteme* 5 (01): 33–55.
- . 2009. *Democratic Legitimacy*. London: Routledge.
- . 2012. "The Procedural Epistemic Value of Deliberation." *Synthese* 190 (7): 1253–66.
- Pettit, Philip. 1996. *The Common Mind: An Essay on Psychology, Society, and Politics*. Oxford: Oxford University Press.
- . 2002. "Defining and Defending Social Holism." In *Rules, Reasons, and Norms: Selected Essays*. Oxford: Clarendon Press.
- Pippin, Robert B. 2008. *Hegel's Practical Philosophy: Rational Agency as Ethical Life*. Cambridge: Cambridge University Press.
- Plato. 1997a. *Crito*. In *Complete Works*, edited by John M. Cooper and D. S. Hutchinson, translated by G.M.A. Grube. Indianapolis: Hackett Publishing.
- . 1997b. *Phaedrus*. In *Complete Works*, edited by John M. Cooper and D. S.

- Hutchinson, translated by Alexander Nehamas and Paul Woodruff. Indianapolis: Hackett Publishing.
- . 1997c. *Republic*. In *Complete Works*, edited by John M. Cooper and D. S. Hutchinson, translated by G.M.A. Grube and C.D.C Reeve. Indianapolis: Hackett Publishing.
- Plutarch. 1960. *The Rise and Fall of Athens: Nine Greek Lives*. Translated by Ian Scott-Kilvert. London: Penguin.
- Putnam, Hilary. 1989. "Reconsideration of Deweyan Democracy, A." *Southern California Law Review* 63: 1671.
- . 2002. *The Collapse of the Fact/value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.
- . 2005. *Ethics without Ontology*. Cambridge, MA: Harvard University Press.
- . 2008. *Jewish Philosophy as a Guide to Life: Rosenzweig, Buber, Lévinas, Wittgenstein*. Bloomington: Indiana University Press.
- Quinton, Anthony. 1975. "The Presidential Address: Social Objects." *Proceedings of the Aristotelian Society*, New Series, 76 (January): 1 – viii.
- Quong, Jonathan. 2010. *Liberalism without Perfection*. Oxford: Oxford University Press.
- Rabe, David. 1995. *Hurlyburly: And, Those the River Keeps : Two Plays*. New York: Grove Press.
- Rawls, John. 1999a. *A Theory of Justice*. Revised Edition. Cambridge, MA: Harvard University Press.
- . 1999b. *Collected Papers*. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press.
- . 2001. *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge, MA: Harvard University Press.
- . 2005. *Political Liberalism*. Expanded Edition. New York: Columbia University Press.
- Raz, Joseph. 1979. *The Authority of Law: Essays on Law and Morality*. Oxford: Oxford University Press.
- . 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- . 1999. *Practical Reason and Norms*. Oxford: Oxford University Press.
- . 2000. *Engaging Reason : On the Theory of Value and Action: On the Theory of Value and Action*. Oxford: Oxford University Press.
- . 2009. "The Problem of Authority: Revisiting the Service Conception." In *Between Authority and Interpretation: On the Theory of Law and Practical Reason*. Oxford: Oxford University Press.
- . 2010. "On Respect, Authority, and Neutrality: A Response." *Ethics* 120 (2): 279–301.
- . 2011. *From Normativity to Responsibility*. Oxford: Oxford University Press.
- Ripstein, Arthur. 2004. "Authority and Coercion." *Philosophy & Public Affairs* 32 (1): 2–35.
- . 2010. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press.
- Rousseau, Jean-Jacques. 1997. *Of the Social Contract*. In *The Social Contract and Other Late Political Writings*, edited and translated by Victor Gourevitch. Cambridge, MA: Cambridge University Press.
- Ryle, Gilbert. 1945. "Knowing How and Knowing That." *Proceedings of the Aristotelian*

- Society*, New Series, 46 (January): 1–16.
- Sandel, Michael. 2005. *Public Philosophy: Essays on Morality in Politics*. Cambridge, MA: Harvard University Press.
- . 2012. *What Money Can't Buy: The Moral Limits of Markets*. London: Penguin.
- Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- . 2003. "Promises and Contracts." In *The Difficulty of Tolerance: Essays in Political Philosophy*. Cambridge: Cambridge University Press.
- Scheffler, Samuel. 1992. *Human Morality*. Oxford: Oxford University Press.
- Schmid, Hans Bernhard. 2009. *Plural Action: Essays in Philosophy and Social Science*. Springer Science & Business Media.
- Schmidtz, David. 1996. "Justifying the State." In *For and Against the State: New Philosophical Readings*, edited by John T. Sanders and Jan Narveson. Vol. 101. London: Rowman & Littlefield.
- Schutz, Alfred. 1967. *The Phenomenology of the Social World*. Translated by Walsh George and Frederick Lehnert. Evanston: Northwestern University Press.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- . 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- . 1995. *The Construction of Social Reality*. London: Penguin.
- . 1998. *Mind, Language And Society: Philosophy In The Real World*. New York: Basic Books.
- . 2003. "Social Ontology and Political Power." In *Socializing Metaphysics: The Nature of Social Reality*, edited by Frederick Schmitt. Lanham: Rowman & Littlefield Publishers.
- . 2005. "What Is an Institution?" *Journal of Institutional Economics* 1 (01): 1–22.
- . 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.
- Sen, Amartya. 1992. *Inequality Reexamined*. Oxford: Clarendon Press.
- Shakespeare, William. 1992. *Hamlet*. In *The Complete Works*, edited by Peter Alexander, New Edition. London: Collins.
- Simmel, Georg. 1910. "How Is Society Possible?" *American Journal of Sociology* 16 (3): 372–91.
- Simmons, A. John. 1981. *Moral Principles and Political Obligations*. Princeton: Princeton University Press.
- . 2001. *Justification and Legitimacy: Essays on Rights and Obligations*. Cambridge: Cambridge University Press.
- Soteriou, Matthew. 2013. *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford: Oxford University Press.
- Stoutland, Frederick. 2002. "Faces of Intention." *Philosophy and Phenomenological Research* 65 (1): 238–41.
- Sugden, Robert. 1993. "Thinking as a Team: Towards an Explanation of Nonselfish Behavior." *Social Philosophy and Policy* 10 (01): 69–89.
- Taylor, Charles. 1985. *Philosophy and the Human Sciences: Philosophical Papers 2*. Cambridge: Cambridge University Press.
- . 1995. *Philosophical Arguments*. Cambridge, MA: Harvard University Press.



- Theunissen, Michael. 1977. *Der Andere: Studien zur Sozialontologie der Gegenwart*. Berlin: Walter de Gruyter.
- . 1984. *The Other: Studies in the Social Ontology of Husserl, Heidegger, Sartre and Buber*. Cambridge, MA: MIT Press.
- Thompson, Michael. 2012. *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press.
- Tocqueville, Alexis. 2003. *Democracy in America*. Translated by Gerald Bevan. London: Penguin.
- Tomasello, Michael. 2010. *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Tuomela, Raimo. 2000. *Cooperation: A Philosophical Study*. Dordrecht: Kluwer Academic Publishers.
- . 2002. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press.
- . 2007. *The Philosophy of Sociality: The Shared Point of View*. Oxford: Oxford University Press.
- . 2013. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press.
- Turner, Mark. 2009. "De Rerum Natura: Dragons of Obliviousness and the Science of Social Ontology." In *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, edited by C. Mantzavinos. Cambridge: Cambridge University Press.
- Unger, Roberto Mangabeira. 1996. *What Should Legal Analysis Become?*. London: Verso.
- Velleman, J. David. 2000. "How to Share an Intention." In *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Waldron, Jeremy. 1999. *Law and Disagreement*. Oxford: Clarendon Press.
- Walker, A. D. M. 1988. "Political Obligation and the Argument from Gratitude." *Philosophy & Public Affairs* 17 (3): 191–211.
- Ware, Owen. 2010. "Fichte's Voluntarism." *European Journal of Philosophy* 18 (2): 262–82.
- Weber, Max. 1946. "Politics as a Vocation." In *From Max Weber: Essays in Sociology*, edited by H.H. Gerth and C. Wright Mills, translated by H.H. Gerth and C. Wright Mills. New York: Oxford University Press.
- . 1978. *Economy and Society: An Outline of Interpretive Sociology*. Edited by Roth Guenther and Claus Wittich. Translated by Ephraim Fischhoff et al. 2 vols. Berkeley: University of California Press.
- Weil, Simone. 2002. *The Need for Roots: Prelude to a Declaration of Duties Towards Mankind*. London: Routledge.
- Williams, Bernard. 1976. "Ethical Consistency." In *Problems of the Self: Philosophical Papers 1956-1972*. Cambridge: Cambridge University Press.
- . 1981. *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.
- Williams, Robert R. 1997. *Hegel's Ethics of Recognition*. Berkeley: University of California Press.
- Winch, Peter. 1972. "Nature and Convention." In *Ethics and Action*. London: Routledge and Kegan Paul.
- Wittgenstein, Ludwig. 2010. *Philosophical Investigations*. Edited by P.M.S Hacker and

- Joachim Schulte. Translated by G.E.M Anscombe, P.M.S Hacker, and Joachim Schulte. Fourth Edition. Oxford: Wiley-Blackwell.
- Wolff, Jonathan. 1995. "Political Obligation, Fairness, and Independence." *Ratio* 8 (1): 87–99.
- Wolff, Robert Paul. 1998. *In Defense of Anarchism*. Berkeley: University of California Press.
- Yaffe, Gideon. 2006. "Reasonableness in the Law and Second-Personal Address." *Loyola of Los Angeles Law Review* 40: 939.
- Young, Iris Marion. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.
- Zimbardo, Philip. 2007a. *The Lucifer Effect: Understanding How Good People Turn Evil*. New York: Random House Publishing Group.
- . 2007b. "Revisiting the Stanford Prison Experiment: A Lesson in the Power of Situation." *Chronicle of Higher Education* 53 (30): B6–7.