**Party to the**

**Hypothetical Contract**

*Obligation, Legitimacy, and Autonomy*

**Carl Fox**

# Thesis submitted for the degree of

# Doctor of Philosophy

# (Philosophy)

# August 2014

# Department of Philosophy

# The University of Sheffield

# *Abstract*

Since Ronald Dworkin rubbished the idea that hypothetical consent could bind an agent in the same way that her actual consent does, social contract theorists have concentrated on refining a purely justificatory strategy for establishing moral and political principles. They have shied away from questions about why, and especially how, we are bound to keep off the grass when a suitably just government tells us to keep off the grass.

I develop a sanction theory of obligation that explains the bindingness of obligations using the cost to an individual of undermining her practical identity. I then explicate the role of hypothetical consent as expanding a practical identity from core principles. To establish an individual’s hypothetical consent to some principle is to invite her to exact a cost from herself in terms of her own self-understanding if she violates it. Although an individual cannot act as her own commander, she can police her commitments. I also argue that our commitment to our practical identities can, and should, be voluntary since we aim to autonomously construct or endorse our identities on the basis of reasons. My goal is to reset the relationship between the social contract and obligation, explain the binding force of obligation, and maintain important roles for hypothetical consent and the ideal of voluntariness within the contract tradition.

Applying this analysis, I recommend new approaches to understanding political legitimacy and what it means for an agent to live an autonomous life. I then go on to outline a strategy to complement the traditional social contract model. Along with determining the kind of state it would make sense for us to agree to support and obey, we should also work out the identity that a citizen must have if she is to bind herself to a just political community. A complete picture of political obligation is one in which the just state is embraced by a conscientious citizen. To achieve this, I propose that we turn the social contract on its head.

**Acknowledgments**

Faced with the prospect of writing a gushing acknowledgments page it is more than a little tempting to say something glib, drop the microphone, and saunter off. But that would be the easy way out. Many people deserve a proper thanks, and that is the least I owe them.

I would like to start by expressing my gratitude to the University of Sheffield for funding my research and to the staff and students of the Department of Philosophy for providing such a supporting and stimulating environment in which to work.

My supervisors Jimmy Lenman and Daniel Viehoff have been a constant source of encouragement, good ideas, and even better questions. Without their patience and guidance this thesis would quite simply not be fit for human consumption. All mistakes, inaccuracies, ambiguities, misinterpretations, invalid arguments, outlandish claims, off-topic ramblings, well… you get the picture. Jimmy and Daniel can take credit for the parts that make sense. I’ll shoulder the blame for the rest.

A very special thanks goes to my girlfriend Jessica Begon who read, and commented on, every single word. I would be a much lesser man and this would be a much poorer thesis without her help. (It would also have far, far less commas.)

I know that I am extremely fortunate to have parents who value education and the pursuit of knowledge so highly. More than that, they are wonderful people and have been far more supportive than any twenty-nine-year-old student has any right to expect. Kevin, Brian, and Lisa less so, but my siblings do have their good points too.

Over the course of my time in Sheffield I have been privileged to know a great many fine philosophers and friends. For slogging through chapters of my dissertation I would like to thank Jessica Begon (again), Jack Wadham, and Joe Saunders. For some good sessions and fun arguments a more than honourable mention goes to Charlotte Alderwick, Heather Arnold, Al Baker-Graham, Josh Black, Peter Caven, Ryan Doran, Rich Healey, Stephen Ingram, Katharine Jenkins, Jan Kandiyali, Natasha McKeever, Armin Khameh, Jonathan Parry, Angie Pepper, Janosch Prinz, Kathy Puddifoot, Philipp Rau, Kim Redgrave, Robin Scaife, Neil William Williams, and Steve Wright.

Over the last year I have had the privilege to get to know Frank McDermott. Sadly Frank passed away recently, but I will not forget our conversations which were a great help to me.

Last and least I would like to thank all of my friends for putting up with me in general. You know who you are, and if you don’t then frankly you’ve been skating along and need to try harder.

Thank you all.

Peace out.

# *Contents*

**Introduction 1**

**Chapter 1 – The Trouble with Hypothetical Consent 7**

*1.1 Introduction* 7

*1.2 The Social Contract* 9

*1.3 Justification, Legitimacy, and Obligation* 11

*1.4 Hypothetical Consent and the Ideal of Voluntariness* 19

*1.5 The Relevance of Consent* 23

*1.6 Autonomy, Anscombe, and the Binding Force of Obligation* 31

*1.7 Conclusion* 34

**Chapter 2 – The Cost of Breaching an Obligation 36**

*2.1 Introduction* 36

*2.2 Social Contract Theory and Obligation* 38

*2.3 Desiderata of a Theory of Moral Obligation* 43

*2.4 Command and Commanders* 46

*2.5 Sanctions and Conscience* 52

*2.6 Intelligibility* 58

*2.7 Voluntariness* 61

*2.8 Conclusion* 67

**Chapter 3 – Sanctions & Self-Interest 69**

*3.1 Introduction* 69

*3.2 Practical Identity* 71

*3.3 The Threat of Self-Interest* 80

*3.4 Sanctions and Exclusionary Reasons* 83

*3.5 Doing the Right Thing for the Right Reasons* 87

*3.6 Grace* 92

*3.7 Conclusion* 95

**Chapter 4 – Legitimacy & Practical Identity 97**

*4.1 Introduction* 97

*4.2 Normative and Descriptive Legitimacy* 100

*4.3 The Foundations of Legitimacy* 104

*4.4 Extending a Practical Identity* 109

*4.5 Authority, Authoritativeness, and Legitimacy* 113

*4.6 Advantages of a Relational Account Tied to Obligation* 118

*4.7 Legitimacy and Voluntariness* 122

*4.8 Democratic Persuasion* 125

*4.9 Conclusion* 128

**Chapter 5 – Autonomy & Living an Intelligible Life 131**

*5.1 Introduction* 131

*5.2 Three Kinds of Autonomy* 134

*5.3 The Value of Autonomy* 139

*5.4 Personal Autonomy* 143

*5.5 Autonomy as Identification with the State* 147

*5.6 Autonomy as Reason-Responsiveness* 153

*5.7 Voluntariness and Living an Intelligible Life* 157

*5.8 Procedural and Substantive Conceptions of Autonomy* 159

*5.9 Conclusion* 161

**Chapter 6 – Inverting the Contract 163**

*6.1 Introduction* 163

*6.2 Rational Reconstruction of the Individual* 166

*6.3 Obligations and Obligatory Ends* 171

*6.4 Stability* 175

*6.5 The Significance of an Overlapping Consensus* 181

*6.6 The Role of Ideals and Ideal Theory* 185

*6.7 Conclusion* 190

**Conclusion 192**

# *Introduction*

Most of us would agree that we ought to keep off the grass when a just political authority tells us to keep off the grass. Indeed, there are many principles to which we would agree if only we were asked, and even more that we would agree to if it were possible to draw down the veil of ignorance and control for the personal ties and biases that sometimes lead us astray in our practical reasoning. But is there any significance to consent that has not actually been given? Indeed, should we care about hypothetical consent at all?

You might well think that we should not. Hypothetical consent is not an approximation of actual consent. In fact, it is no consent at all.[[1]](#footnote-1) Arguments must stand or fall independently of what people would agree to if asked. All the same, in this dissertation I want to push back against the widespread notion that hypothetical consent is good for nothing more than window-dressing on any theory that seeks to ground our political obligations. However, I also want to avoid collapsing into a view that rescues hypothetical consent at the cost of the voluntarist ideal that animates the social contract tradition. On this view, the hypothetical consent of representative parties, when conceived very carefully under specific conditions, provides the standard for the moral justification of principles to which we are thus bound. There is no need for voluntary participation on the part of real citizens in the generation of their duty to obey the commands of a suitably just state.

On my preferred view, hypothetical consent can also perform an indispensable theoretical and political function by extending the practical identities of prospective citizens from principles that they already endorse to further principles and ends that can be shown to follow from their original commitments. I will, therefore, argue that the value of hypothetical consent is neither merely illustrative nor purely justificatory.

Treading this fine line leads me to propose an original theory of the binding force of obligation. To be under an obligation is, I think, ultimately to relate to the reasons that support it in a special way. It is to conceive of oneself through those reasons and thereby to incorporate them into one’s identity. Violating a principle to which you are committed in this way comes at a high price. It undermines and potentially even unravels your self-conception. This is something that almost all of us want to avoid and I will argue that it constitutes an appropriate sanction to underpin a theory of the bindingness that is phenomenologically characteristic of obligation.

A consequence of this view is that an obligation only exists for a person if there is the right fit between the reasons that support it and how she thinks about herself. This does not affect our right to punish or restrain those individuals who commit moral wrongs. In fact, it points us towards the importance of identity and identity formation in securing the influence of those considerations that should regulate our moral and political behaviour.

There are a number of advantages to adopting the account of obligation I defend here. As I will show, it creates conceptual space for the possibility of grace in those rare moments when we do our duty cheerfully and spontaneously. It also explains how we can do things for the right reasons even when we feel the weight of duty pressing down upon us. It facilitates new perspectives on both legitimacy

and autonomy, and it can be applied to shed light on what it means for a theory of justice to be stable and on issues in the debate between ideal and non-ideal theorists.

I will begin in Chapter 1 by considering the role of hypothetical consent in a contract theory broadly conforming to the mould set by John Rawls and move on to explore the relationship between justification, legitimacy, and obligation. Two broad views will be examined. One holds that no amount of theorising can establish an obligation to do what the state commands because it is only individuals who have the normative power to subject themselves to a political authority. This will be contrasted with the idea that sufficient justification can determine binding moral principles, including the duty to obey just states. I will argue that although the first view cannot stand, it is a mistake to circumvent the ideal of voluntariness and to endorse an account of political obligation that does not include a role for the participation of individuals in the generation of the binding force of their own obligations. Any attempt to ground obligations must answer a general challenge about the intelligibility of the concept laid down by Elizabeth Anscombe. I will show why this is a particular concern for hypothetical consent theorists before briefly sketching my own answer that I will defend in subsequent chapters.

Chapter 2 duly tackles obligation head-on and develops an original theory of the binding force of obligation that relies on the sanction of unintelligibility that follows when an action is incompatible with an individual’s self-understanding. Although Anscombe is right to deny that individuals can act as their own commanders, they can still police their own commitments by exacting a cost in the event of a breach. I shall defend and adapt Serena Olsaretti’s account of voluntariness, on which an action counts as voluntary if it is not performed because of the unacceptability of the alternatives. The existence of a sanction explains the unacceptability of breaching an obligation but it is also a sanction that can be imposed voluntarily since it is possible to willingly embrace one’s moral principles.

Of course, there are many objections to sanction theories of obligation and Chapter 3 will deal with the most troubling worries and then proceed to further elaborate on the details of my own theory. In particular, implicating self-interest might be thought to produce compliance for the wrong reasons. It might also be objected that sanctions can be evaded and that an identity-based view creates a perverse incentive to change one’s self-conception as a means of escaping the cost of violating a duty. I will address these worries by showing that although we must care about living an intelligible life, this is the only egoistic prerequisite for being subject to the binding force of obligation. With some assistance from Christine Korsgaard’s notion of a practical identity, I will argue that my view is uniquely equipped to explain moral continence, and even how it is possible to do our duty cheerfully.

Having outlined and defended my account of the binding force of obligation, the challenge will then be to explore its implications for other concepts that are tied to it in a more complete contract theory. Chapter 4 recasts political legitimacy accordingly. My argument will be that political legitimacy is not the same as sufficient moral justification to ground a state’s right to wield coercive power. A state is, rather, legitimate with regard to particular citizens on a case-by-case basis insofar as the justification for its authority fits with, and follows from, key elements of their practical identities. Only when this relationship exists can a state deploy the binding force of obligation by extending citizens’ identities around new principles, policies, and laws. Further, when this relationship obtains a state can play a meaning-giving role by establishing and protecting structures and institutions that we have reason to support. In this way a state can enhance individual autonomy by helping us to live intelligible lives.

What it means to live an intelligible life provides the basis for the account of autonomy that I will present in Chapter 5. We have, I believe, a deep interest in being reason-responsive and my main contention will be that it is only by binding ourselves to a hierarchy of principles that our responses to reasons combine to create a unique, concrete individual identity. We need, therefore, a substantive account of the basis of autonomy that goes beyond the procedural conditions necessary to be sensitive to normative considerations. I will also argue that not only is autonomy consistent with obedience to a suitably constituted authority, it can be greatly enhanced by it. Two of the most important ways that states serve our interests is to assist us in responding to reasons by making our commitments more determinate and by facilitating a collective response to some problems that are simply too large for any individual to surmount on her own.

Finally, in Chapter 6 I will suggest a strategy to complement the contract approach that employs the ideas developed in earlier chapters. The traditional model starts with a particular conception of the person and attempts to work out terms it would make sense for such a person to agree. I propose that contract theorists should also build the matching identity that a citizen must have if she is to bind herself to a just political community. I will refer to this strategy as ‘inverting the contract’ because it shifts the focus to what an individual will have to be like if she is to relate to the conclusions of a traditional contract theory in an obligatory way. In order to prove the usefulness of this proposal I will spell out how it applies to the question of the stability of a conception of justice and use it to shed some new light on the ideal/non-ideal theory debate. In the first case, I argue that stability can only be achieved when citizens incorporate the specific demands of a conception of justice into their self-conceptions and, via the self-imposed sanction of unintelligibility, secure for those demands a regulative role in their deliberations. In the second, my contention will be that ideal theory sets important constraints on the compromises that non-ideal theory can make to challenging circumstances. States should not implement policies that are incompatible with their responsibility to promote identities in their citizens that would make progress towards an ideal end-state obligatory.

So, what is the significance of all of this? There are two overarching points to take away. The first is that relating to a principle in an obligatory fashion is an elegant and effective means of responding to reasons. By voluntarily binding herself to principles, a person shapes her own self-conception into a framework for living an intelligible life. The binding force of obligation is woven into the way that we understand our own selves. This, I think, is why the concept of obligation is so deeply embedded in our political and moral discourse and why it would be a terrible shame to give up on it.

The second point is that there is life in the social contract yet. More particularly, hypothetical consent can be picked up, dusted off, and put to, if you’ll pardon a pun at this early stage, actual use. We are interested in justifying moral and political principles as reasoning beings who engage in the world over the course of a life*.* Hypothetical consent arguments are not purely illustrative because they can tell us something that will matter to us in this regard: how we must relate to their conclusions if we are to remain intelligible to ourselves. The following chapters will, I hope, clarify and develop these ideas.

# 1

# *The Trouble with Hypothetical Consent*

**1.1 – Introduction**

When Hobbes attempted to establish that we *ought* to obey the Leviathan, he tied the success of the social contract to the generation of the most basic political obligation: obedience to whoever is in charge.[[2]](#footnote-2) Since then, the ability of contract theories to ground obligations has been the focus of varied, but relentless, criticism. In this dissertation I intend to defend contract views that deploy hypothetical consent on the basis that hypothetical consent can play a key role in the generation of the *binding force* of obligation.

My jumping-off point is an important debate about the relationship between justification, legitimacy, and obligation in modern contract theory. A. John Simmons has argued that even though a hypothetical social contract argument might be able to successfully justify some set of moral or political principles, it cannot, by itself, establish that we are bound to pay our taxes when a suitably just government tells us to pay our taxes. The problem is that no amount of justification is equivalent to a voluntary action on the part of some agent that she understands will constitute submission to an authority. Cynthia Stark has responded by arguing that hypothetical consent is not intended as a poor substitute for actual consent. Rather, it is the standard of legitimacy for fundamental moral principles, to which actual consent is not relevant. Nobody thinks that we need to agree not to murder for it to be the case that we are bound not to murder. Stark argues that the same is ultimately true of other duties, such as the duty to obey just states.[[3]](#footnote-3) To establish the hypothetical consent of representative parties under certain carefully specified conditions just is to establish our basic moral principles and it is on this basis that they exert binding force upon us.

There are three problems with this solution. First, if our actual consent is not relevant to fundamental moral obligations there is no cause to think of them as voluntary or in any significant way self-imposed. This makes a nonsense of the deference shown to the ideal of voluntariness throughout the contract tradition. Second, Stark’s response fails to capture the robust intuition that many of us have that we must play some part in generating our political obligations. By relying on the justification of foundational moral principles she appears to bundle them in through the back door and that seems too easy. Third, it opens her up to Anscombe’s general worry about the bindingness of moral principles in the absence of a divine commander. In what sense do we *have* to comply with the decisions made by our idealised representatives?

The social contract, it seems, has a problem with obligation. How does it coherently generate a binding duty of obedience to a suitably constituted political authority? In this chapter I will motivate the puzzle of obligation for the contract theorist by exploring the relationship between legitimacy and obligation, and then

between obligation and autonomy. Later chapters will pick up the various threads and develop these concepts in much more detail.

I will begin by discussing the idea of the social contract and specifying the version that I will focus on defending here.[[4]](#footnote-4) Section 1.3 will turn to the relationship between justification, legitimacy, and obligation and set the stage for the clash between Simmons and Stark. The next section will then consider the significance of hypothetical consent as the source of moral principles and what that means for the ideal of voluntariness before I raise the question of the relevance of consent to our duty of obedience to a just state in Section 1.5. Respect for autonomy and Anscombe’s objection to the Kantian picture of obligation as self-legislation will be the subject of Section 1.6, before I conclude and briefly sketch my intentions for future chapters.

**1.2 – The Social Contract**

The core feature of any contract view is that it deploys the ideal of agreement to establish that we have reason to accept some principle or some set of principles. I am not interested in theories that employ the idea of agreement only as a heuristic, to illustrate an independent argument. Agreement must play some integral role in the whole theory.[[5]](#footnote-5)

Contract theories must also specify who is to be party to the agreement. Not only can this range over different numbers or groups of people, it can extend to theoretical constructs who represent our interests. This move permits contract theorists to eliminate the effect of arbitrary information, by drawing down the so-called ‘veil of ignorance’ over the parties, and to idealise some of their capacities so that they are not derailed by faulty reasoning.[[6]](#footnote-6) I will be concerned here primarily with hypothetical consent theories where the objective is to determine what suitably specified representative parties *would* agree to.[[7]](#footnote-7) Of course, it is also crucial to determine the purpose of the contract; what question, or questions, are we using it to answer?

As I will describe in this chapter, there has been a move away from the traditional concern of Hobbes, Locke, and Rousseau with anchoring the duty of obedience to a legitimate political authority. This move was prompted by the failure of hypothetical consent to generate obligations in the same way as our actual consent.[[8]](#footnote-8) Contract theorists as different as Samuel Freeman (1990) and David Gauthier (1986) focus instead on providing justifications. In Freeman’s (1990, p.128) case it is for a system of norms “for assessing individuals’ ends and desires, and the courses of conduct they adopt in order to realize them”, while Gauthier (1986, p.20) seeks to defend the possibility of morality by forging “a link between the rationality of individual maximization and the morality of impartial constraint”. Similarly, we could seek to demonstrate that some kind of state would be worthy of having a right to rule, or recommend a set of rules to regulate a more particular sphere of human interaction,[[9]](#footnote-9) etc… On this line, hypothetical consent provides a standard of justification that the conclusions of the argument then must meet. The source of our obligations is simply a different matter.

This, I think, is a mistake. There is a deep connection between the justification of political authority and the generation of an obligation to obey it. This is tied up in two other ideas that are central to the contract tradition, legitimacy and autonomy, which I will explore in Chapters 4 and 5 respectively. For now, I will begin in earnest by teasing out the implications of a puzzle about how justification might be used to ground moral obligations and thereby, through the back door, the political duty of obedience. If obligation really is beyond the reach of the social contract then the move to a purely justificatory approach would be a necessary one.

**1.3 – Justification, Legitimacy, and Obligation**

At the heart of this puzzle are two conflicting views about the possibility of establishing political obligations from the armchair, so to speak, and what follows from the justification of political authority. On one view, we can theorise our way to a full justification of moral principles that bind individuals to obey a just state and authorise such a state to deploy coercive force. ‘Legitimacy’ covers both a right to rule and the correlative obligation to obey. To be a legitimate authority is to have the right to issue, hold, and enforce binding obligations. Hanna Pitkin (1966, p.39) sums up this thesis succinctly, saying that “legitimate authority is precisely that which ought to be obeyed”, and, therefore, “your obligation is to obey what deserves obedience and consent” (1966, p.42). I shall call this position the ‘sufficiency thesis’ because it holds that to justify political authority just is to establish that we are bound to obey it.[[10]](#footnote-10)

However, Simmons, taking his cue from Locke, objects that political authority must first be justified against a background assumption of liberty by demonstrating that it is both morally permissible and superior to a state of anarchy. If both of these conditions can be met then some kinds of possible authorities would be justified, but this is not the same thing as to say that any particular person ought to obey any particular state. Rather, the advisability of a coercive state and the commitment to obey it “are simply independent values, in the same way that the right of some business to provide services to you and to bill you for them is independent of that business’s efficiency or generosity or usefulness” (Simmons 1999, p.752).

It is worth noting that Simmons also uses ‘legitimacy’ to denote the political power to create and enforce binding obligations. However, he drives a wedge between the justification of authority and the generation of a duty of obedience by advocating a principle that insists that a duty of obedience to a political authority depends on some prior voluntary action that is understood to create this duty.[[11]](#footnote-11) This might be the explicit or implicit purpose of the action (or meaningful inaction)[[12]](#footnote-12) in cases of consent, or an acknowledged consequence of accepting and enjoying the benefits provided by a scheme of cooperation.[[13]](#footnote-13) Simmons, therefore, distinguishes between qualifying for the right to wield coercive power by meeting the relevant standard of justification[[14]](#footnote-14) and acquiring the authorisation that may be required over and above that in order to permissibly give orders and back them up with force. A general standard of justification is still useful on this view because it defines the range of suitable candidates for important and morally significant roles.

We can characterise this position as the ‘insufficiency thesis’. It states that justification alone is insufficient to ground political obligations and it captures something important in our evaluation of political authorities: to say that a state is legitimate is usually to imply something beyond the general case that can be made for it, something more particular about why *we* are under an obligation to do as we are told.

The sufficiency thesis stands in stark contrast since it holds that justification is, ultimately, sufficient to ground our political obligations. Why would anyone recommend it? Well, if we assume with Rawls (2003, p.4) that “we simply find ourselves in a particular political society at a certain moment of historical time” and have no initial choice over the basic structure of that society, “its main political and social institutions and the way they hang together as one system of cooperation” (2003, p.8), then it appears fruitless to insist on wringing authorisation for a coercive state from the voluntary acquiescence of citizens.[[15]](#footnote-15)

We might instead concentrate on a seemingly different task, designing a fair and equitable basic structure which will, after all, “affect men’s initial chances in life” (Rawls 1999, p.7). Whether or not somebody, or some body, is entitled to tell us to keep off the grass is a secondary concern, although still very much a relevant one. For a Rawlsian contractualist, we proceed by discerning what an idealised group of representative parties would agree to when we abstract them away from the personalised cares and concerns that would introduce morally arbitrary features to the equation and corrupt their deliberations.[[16]](#footnote-16) The point of all of this is to find the basic principles to which it would be reasonable to expect people to give their consent. In this way, figuring out what it is that free and equal agents prepared to engage with others on fair terms would agree to, we can determine just principles to regulate and secure social cooperation. This is not all, however. This process is how we come to see what the justification of fundamental principles actually means, and that is showing what it is that we, as a community of reasonable people, can expect individuals to endorse as reasonable.

There are different ways of spelling this out. Thomas Scanlon (1998), for instance, builds his theory around what individuals cannot reasonably reject rather than what we can expect them to endorse. I will not consider the merits of the competing approaches here because I am more interested in how such theories can ground duties of obedience without recourse to the voluntary participation of citizens.

Rawls (2005, p.137) states that “our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason”. He (2005, p.137, fn.5) explicitly argues that we can make sense of this liberal principle of legitimacy by looking at it “from the point of view of the original position”, and that it has “the same basis as the substantive principles of justice” (2005, p.225). In essence, then, what makes the use of coercive force permissible is that it can be justified in the same way as our other fundamental principles.

So, “the task of discovering the conditions of legitimacy is traditionally conceived as that of finding a way to justify a political system to everyone who is required to live under it” (Nagel 1991, p.33). What this means is that it might be possible to ascertain that we in fact have political obligations by determining the principles to which we can reasonably expect people to adhere. This is the basis of the sufficiency thesis. It turns out that one of the moral principles to which we can be expected to agree mandates obedience to a suitable political authority. I will explain this kind of contract view in more detail in the next section, but before I do I want to briefly consider some initial reasons for suspecting that the sufficiency thesis is flawed and to draw out the general response that sufficiency theorists make to the concerns raised by the insufficiency thesis.

One such worry is that we seem to be able to conceptually separate legitimate authority from obligation, and, indeed, obligation from legitimate authority. The paradigm case is foreign states whose legitimacy we recognise, and perhaps endorse, but whose laws and decrees are not binding outside of their jurisdiction. As an Irish citizen, I can hold that the British state is legitimate. However, its legitimacy does not entail that I should obey its laws in Ireland, nor anywhere else in the E.U., in international waters, or even if I find myself in a failed state such as Somalia. The brute fact of its legitimacy is insufficient to generate global obligations of obedience towards it. This is commonly referred to as the ‘particularity requirement’ and it suggests that some element of our personal or individual circumstances must be relevant to the origin of obligations to legitimate authorities.[[17]](#footnote-17)

It looks like we can also break apart the relationship from the opposite direction by showing that it is possible to have obligations to obey illegitimate states. Say that there is a catastrophic natural disaster in North Korea and you go as a member of a humanitarian mission to help the injured and the displaced. On arrival it turns out that the North Koreans intend to take full charge of all rescue and resettlement operations and they order you to distribute your aid by certain methods and employ your skills in certain ways. Most of us would agree that there is an obligation to help the stricken people, and it seems plausible that it may actually give rise to an obligation to obey the dictates of a corrupt and inefficient government. Further, you may explicitly consent to ‘play by their rules’ in order to help the unfortunate victims and thereby come to feel bound to obey their commands. It would seem very odd indeed to argue that either the circumstances or your agreement has transformed an oppressive regime into a legitimate government, but a reasonable case can be made that you have acquired a limited set of specific obligations to it. At the very least, it would be natural to *feel* that your consent commits you, in the way that consent usually does.[[18]](#footnote-18)

The obvious response to the latter objection is that we simply cannot have obligations to obey illegitimate authorities because they are disqualified, either on moral or prudential grounds. Rawls (1999, p.96) argues that “it is not possible to be have an obligation to autocratic and arbitrary forms of government” because they are unjust, while both Locke (2004, p. 284) and Simmons (1979, p.87) suggest that the practical purpose of political obligations preclude us from owing them to entities such as tyrannical governments or failed states since they cannot serve our interests.[[19]](#footnote-19) This means that a promise to obey the government of North Korea cannot generate an obligation to recognise its legal system, serve in its army, or pay parking fines to its traffic police.[[20]](#footnote-20)

Rawls heads off the particularity problem with his natural duty of justice that “requires us to support and to comply with just institutions that exist and *apply to us*” (1999, p.99, my emphasis). But what can Rawls mean by ‘apply to us’?[[21]](#footnote-21) Jeremy Waldron (1993, p.13) builds an account around the idea that we can distinguish between “insiders” and “outsiders” with regard to the domain of principles of justice. Such principles are limited in their range because they are designed to apply to discrete localised groups for the reason that they are the first site of conflict between individuals in a state of nature.[[22]](#footnote-22) Having argued this, he thinks that we can hold a natural duty of justice to be binding if we can satisfy the following three conditions: (1) show that it is important for there to be institutions ‘doing justice’; (2) demonstrate that there should be only one such institution in a particular area; and (3) identify grounds for favouring one particular organisation over others. If we satisfy these conditions then “an organisation that is just, effective, and legitimate (in the sense of being singled out as *the* salient organization for this territory) has *eo ipso* a claim on our allegiance” (1993, p.27, original emphasis).

Waldron holds that “it is morally imperative that the demands of justice be pursued *period*” (1993, p.28, original emphasis), but this merely asserts the sufficiency thesis and obscures a deeper worry about contract views. In the broadly Rawlsian contract, the hypothetical consent of the representative parties is the standard of justification which generates our commitment to justice. But how can we be bound to the terms of a contract that nobody signed? Simmons (1999, p.752) argues that “only interacting with you – and in a way that we normally suppose gives one party a moral right to expect something of another – will seem to ‘legitimate’ its imposition and/or enforcement of duties on you”.

The sufficiency theorist can only escape out from under the particularity problem if we grant his major premise and concede that legitimacy can be established by justification alone. Thus Thomas Nagel (1991, p.36) says that “the search for legitimacy can be thought of as an attempt to realize some of the values of voluntary participation, in a system of institutions that is unavoidably compulsory”, and Rawls (1999, p.12) agrees: “a society satisfying the principles of justice comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are autonomous and the obligations they recognise are self-imposed”. The question now arises, does this nod to the value of voluntary participation via respect for autonomy mollify the voluntarist intuitions that motivate the insufficiency thesis?

**1.4 – Hypothetical Consent and the Ideal of Voluntariness**

Stark (2000) takes a radically different approach, arguing that hypothetical consent theorists like Rawls and Nagel are actually compelled to drop their voluntarist rhetoric on pain of incoherence. The reason for this harks back to what she (2000, p.317) refers to as the “standard indictment” of hypothetical consent.

Ronald Dworkin (1975, p.17-21) famously argued that hypothetical consent is no consent at all and, therefore, cannot be used as a substitute for actual consent. Regardless of whether it could be established that anybody *would* give their consent to some potential contract or agreement, Dworkin maintains that in the absence of their actual consent, we simply have no permission to enforce such a contract. Hypothetical consent provides no argument beyond the merits of the contract itself, and that is insufficient to bind anybody in the way that actual consent normally does. If, for example, I would have paid £10 to have my hair cut, this does not grant licence to a philosophically-minded guerrilla barber to do the job and then demand my £10 afterwards.

At this point Stark (2000, p.318) makes a vital move, noting that “critics of hypothetical-consent theories are not concerned with the problem of the bindingness of moral rules generally, but rather with the bindingness of those principles to which one might think consent is especially relevant”. What she means here is that we think that our actual consent is important for the generation of certain types of obligations, but not others. My agreement is required if I am to be bound to proof-read my friend’s thesis, but not to my obligation to refrain from murdering people. Critics of hypothetical consent theories generally address their concerns to obligations where it seems that actual consent is particularly relevant and they accuse their opponents of trying to bind people in these instances with no consent at all.

Stark goes on to draw an intriguing distinction between political and moral legitimacy. She contends that contract theories can be divided into two categories based on a corresponding division of labour. The position that she calls ‘moral contractarianism’ is concerned with the justification of our basic moral principles.[[23]](#footnote-23) It is employed to identify our commitment to principles that have nothing to do with our actual consent. As such, the use of hypothetical consent in their justification is not intended to mimic the operation of actual consent and so these theories are not troubled by the standard indictment.

She makes the case that hypothetical consent in moral contractarianism is intended to give us reasons for abiding by foundational moral rules and it is inappropriate to adopt the same approach for properly political principles. Good reasons do not bind us when consent is especially relevant because actual consent is the only normative power that can establish an obligation to proof read a friend’s thesis or pick her up from the airport. So, the fact that I would agree to pay taxes towards the upkeep of a good and just government cannot generate an obligation to pay those taxes. Political contractarianism, Stark argues, tries to accomplish an impossible task by grounding obligations to which consent is relevant in hypothetical consent and is subsequently defeated by the standard indictment.

Stark (2000, p.333) groups Rawls and Nagel under the moral contractarian banner and then makes a neat move on their behalf to establish the coercive authority that eludes the political contractarian. Her solution is that they justify the enforcement of apparently political principles by virtue of a higher-order moral principle. The natural duty of justice to obey and support just states that Rawls (1999, p.99) identifies is backed up by his argument that it “would be chosen in the original position; it is a moral principle, as opposed to a political principle, justified by means of hypothetical consent” (Stark 2000, p.331). What this means is that Rawls doesn’t need any actual consent to bind individuals to obey and support just states. Even though we tend to think that our participation is relevant, our moral duties in fact supersede any requirement for actual consent so hypothetical consent can ground the authority of just states to coercively enforce their laws. If there was no duty to obey just states then political authority would depend on consent, but this is precisely why Rawls sets the natural duty of justice over the duty of fair play. The latter cannot adequately support just political authorities because it is ultimately grounded in the free actions of individuals.[[24]](#footnote-24)

Couched in the language that we used to describe the sufficiency thesis; justifying principles by showing that idealised negotiators in certain conditions would agree to them shows both that they are legitimate and that we are bound to comply with them. So far, so good we might think. However, this turns out to be a costly move for a contract theorist. For one thing, it looks like Stark (2000, p.334) is right to insist that the deference paid to the ideal of voluntariness should be dropped: “the fact that the source of authority of political principles is, in hypothetical-consent theory, located in citizens’ rational willing does not make citizens compliance to those principles voluntary, if those citizens are not in fact willing to obey and are, moreover, forced to obey by an external authority”. Indeed, Rawls himself (1999, p.294) appears to admit as much when he says that “each is bound irrespective of his voluntary acts, performative or otherwise”.

It is still important that individuals retain control over many aspects of their lives but on this scheme “questions of actual consent arise only as internal questions of liberty, that is, as questions about what options acceptable institutions must leave open to those living under them” (Scanlon 1976, p.17). The question is whether it still makes sense for Nagel (1991, p.36) to insist that “the search forlegitimacy can be thought of as an attempt to realize some of the values of voluntary participation, in a system of institutions that is unavoidably compulsory”.[[25]](#footnote-25)

There is certainly a concern for autonomy. Rawls (1999, p.455) is clear that: “We are not literally to respect the conscience of an individual. Rather we are to respect him as a person... There is no violation of our autonomy so long as its principles are properly followed”. But we need to dig deeper and inquire about the relationship between autonomy and obligation. In fact, when we get to grips with obligation in Chapter 2 we will be able to see that an individual’s participation is essential to the existence of a genuine requirement since she must police her commitments. Moreover, that participation can be voluntary and I will explain in Chapter 5 why voluntarily binding ourselves to those principles which merit our commitment is an important expression of autonomy. In this chapter, however, I will concentrate on a related question which is how a sufficiency theorist understands fundamental moral duties to be autonomously imposed. This will be the subject of Section 1.6. Before that I will take a closer look at the merits of the insufficiency thesis because its stubborn insistence on the necessity of voluntary participation points to the significance of how we relate to our binding moral commitments.

**1.5 – The Relevance of Consent**

Another way to come at the question of the relevance of consent is to focus on the source of the voluntarist’s unease with attempts to tie individuals to principles via their idealised counterparts, the ‘back-door’ aspect of the political obligations that Stark derives from our fundamental moral duties. Simmons (1999, p.62-64) is prepared to grant that there are natural laws in the sense of moral rules that apply to us independently of our acquiescence, but for him obligations to obey political entities are not derived from this source. Stark’s explanation for why political contractarians cannot use hypothetical consent to determine a contract that establishes obligations to obey civil laws is that consent is relevant to these commitments in a way that it isn’t to the basic tenets of morality. Yet, it would seem that we find ourselves bound to exactly these troublesome, consent-relevant obligations through the mechanism of hypothetical consent, albeit at a deeper level. It is not difficult to see why this would breeds suspicion that a sleight of hand is taking place here. If consent is relevant to an obligation then shouldn’t it stay relevant?

A virtue of the insufficiency thesis, indeed its chief virtue, is that it provides an explanation of this intuition by appealing to a normative mechanism of obligation that stands apart from the reasons that we might have for endorsing some institution or principle. Locke, for example, points to the actual consent[[26]](#footnote-26) of the citizenry as the condition that must obtain for the use of coercive political power to be permissible, no matter how sensible or desirable it may be.[[27]](#footnote-27) It is, of course, compatible with this thesis that demonstrating that some variety of state is morally permissible could be a necessary step in establishing any obligation to obey it,[[28]](#footnote-28) but the reasoning that might establish this cannot be sufficient to generate duties. The key ingredient is voluntary action on the part of the agent who is to hold the obligation.

But how plausible is this view? As Dudley Knowles (2010, p.19) points out, the capacity to effectively wield coercive power is the “nasty face of the state”, because states “threaten their citizens, fine, imprison, publicly shame and exact compulsory service from them. In some jurisdictions they inflict corporal punishment and the death penalty”. It seems, then, that we must think of the state as a necessarily coercive enterprise because the permissibility of all of these social tools is derived from the general permissibility to coerce people in the first place. If such an entity can be shown to be both morally permissible and advisable, then perhaps the same strategy will also show its coercion of unwilling citizens to be morally permissible, in which case why should we worry about whether anyone has a duty to do as it says?

One reason might be that those of us who meet certain procedural standards count as autonomous persons and moral equals, and how can one equal tell another equal what to do? This is not simply a worry about the permissibility of one agent bending another to her will. Insufficiency theorists do not think that it is impermissible to forcibly restrain dangerous criminals, for example. Rather, the concern is with elevating some particular group of individuals and conferring upon them the special ability to issue and enforce binding duties on us. The power of consent provides a solution that respects everyone’s autonomy and equal status. If individuals place themselves under an obligation to obey a state then its authority can always be traced back to a choice they made as autonomous agents.

This is why Simmons (1987) counts himself as a philosophical anarchist. Legitimate authority is possible on his view, but extremely unlikely, and certainly not an accurate description of any existing state. He does not hold to the stronger strain which relies on a conception of autonomy that makes it incompatible with the idea of any authority other than one’s own.[[29]](#footnote-29) Such theories reject the possibility of legitimate states altogether on the grounds that nobody is morally permitted to accept the political authority of any person or institution. Weak philosophical anarchism suggests strongly that the notion of a legitimate authority holding and generating obligations, while coherent and logically possible, is a practical and pragmatic impossibility. Ultimately, this gives us reason to adjust our view of the relevance of consent to our duty of obedience to just states that exist and apply to us since there are powerful moral reasons that support the authority of just states.[[30]](#footnote-30) There is, however, something right in the insufficiency thesis and I will try to extricate it here by considering two interesting theories that attempt to skirt the conflict I have been describing.

In pursuit of an account of legitimacy that does not rely on a correlative obligation to obey, Robert Ladenson (1980) distinguishes between claim-rights and justification-rights[[31]](#footnote-31) within the context of defending a Hobbesian conception of law and sovereignty. “To assert a claim-right is to press a demand with respect to something or other against another individual” (Ladenson 1980, p.137). An important point to note about claim rights is that Ladenson thinks that they presuppose an institutional background of rules and authoritative procedures for enacting and implementing those rules. Claim-rights also “necessarily correlate with duties because when one asserts such a right under the appropriate rules, one makes a claim to certain performances or forebearances by others” (Ladenson 1980, p.138).

Justification-rights, however, are different in that they do not comprise a claim against anyone else. Rather, they can be characterised as sufficient justification for performing some action. They do not correlate with any duties. Ladenson points to such examples as self-defence, defence of others, and parental authority. These are intended to be paradigmatic examples where an agent is right to act, and has a right to act, irrespective of what anyone else is under an obligation to do. He holds that the right to make and enforce laws is a justification right, and, therefore, does not depend on the subjects or citizens of some government being obliged to obey it. A state can be successfully justified if “one can set out a plausible account of a line of reasoning that would lead all rational people under the veil of ignorance not to object to coercion when genuinely carried out by governmental authority” (Ladenson 1980, p.140). There is simply no need for anyone to hold a duty to obey so long as the authority can make the relevant laws, institutions, or principles stick.

The difficulty here, though, is that if there is not sufficient cause to object to some authority wielding coercive force then, if an authority can be shown to apply to us and to meet Waldron’s additional conditions,[[32]](#footnote-32) why should we not think that we *ought* to obey it? A justified authority will be much more effective if we treat it as an authority.[[33]](#footnote-33) In fact, it will also be much more efficient at helping us to conform to our own reasons if we treat it as an authority.[[34]](#footnote-34)

William Edmundson (1998) offers a more conciliatory route to this conclusion. He suggests that we can reconcile the tension between the sufficiency and insufficiency theses by turning down the intensity and scope of the duties that stem from legitimate authority. Specifically, he (1998, p.44) argues that instead of a general duty to obey its laws; “there is a general *prima facie* duty not to interfere with the bona fide administration of the laws of a just state”.[[35]](#footnote-35) The key difference here is supposed to be that where the law is opaque, and its cumbersome language far removed from everyday life, the actual administration of the law is something familiar to all of us, and, crucially, it is something that carries considerable intuitive force. I might feel under no real obligation to obey red lights and stop signs on deserted streets at three in the morning, but when I am confronted by a police officer the situation changes instantly because of the directness of the relationship that now obtains between me and the flesh and blood representative of the state waving me onto the hard shoulder.

This, we are told, makes the duty to obey the “administrative prerogatives” of a just state content independent in a way that its laws simply are not. It is open to us to scrutinise the content of laws before deciding whether or not to obey them, but we must simply do as we are told by the agents of the law.[[36]](#footnote-36) This duty to obey “applicable administrative prerogatives whenever in the circumstances they are entailed by the general prima facie duty not to interfere with their enforcement” (Edmundson 1998, p.54), is therefore built on a more general duty not to undermine or disrupt the enforcement of the laws of a legitimate authority. “Even so, the scope of the duty to which this view is committed is drastically narrower and less objectionable than that of political obligation as traditionally conceived” (Edmundson 1998, p.54), that is, I take it, insofar as we are not under an obligation to obey the laws of a legitimate state just because it is legitimate. Rather, we are perfectly entitled to deliberate on the content of any particular law before deciding to comply with it or not, but we are under a *prima facie* obligation to obey the duly appointed or elected officials of the state.

There are a number of difficulties with Edmundson’s argument, but I will pursue one here that is especially relevant to my concerns in this chapter. Edmundson is clearly interested in telling a story about our obligations to obey legitimate states in which our personal interactions with the state matter. This is why he attempts to explicate political obligations in terms of the relationships that exist between the members of a society and the agents of the state in which they live. While he is perfectly correct to identify these relationships as something that matters to us, he does not explain what role they ultimately play, if any, in establishing or grounding any obligations to obey the agents of a legitimate state. Instead, it seems that this is indeed achieved by the justification backing up the state. The only difference is that we are to find our political obligations somehow less mysterious and onerous when they are delivered to us with, literally, a human face.

Of course, this means that Edmundson sides with the sufficiency thesis, even though he tries to maintain a concern for the personal interactions that characterise our experience of political authority. He fails to defuse the charge of the insufficiency theorist that it is some vital element of our interactions that gives rise to our obligations because, on his view, the obligations are only cashed out in terms of our relationships to power. They are still determined by the legitimacy, or not, of the authority in question. However, there is something to be said for his approach. So, does his claim stand that the binding force of political obligations is less troubling if restricted to administrative prerogatives?

Unfortunately for Edmundson, I do not think so. He has incorrectly identified the sticking point as the *strength* of the political obligations that can be established by justification. The concern, though, is really about the *source* of these obligations and this is motivated by a conviction that obligations like the duty to obey the agents of a legitimate state can only properly be established by a “personal transaction” (Simmons 1979, p.156) with the state such as the granting of one’s actual consent or the voluntary receipt of significant benefits. If we are to adequately respond to Simmons and the proponents of the insufficiency thesis we still need to reconcile the pull of the ideal of voluntariness with the generation of our political obligations.

Edmundson mistakenly emphasises the relationships that obtain between individual citizens and the agents of the state. What matters is actually how we relate to the justification of political authority and to the content of the directives of an acceptable state that exists and applies to us. The insufficiency thesis gets this much right: justification on its own is inert. Simmons is wrong, however, to conclude both that this insight applies only to political rather than moral principles, and that the solution is to rely on the special normative power of actual consent. I will go on to argue that we are genuinely bound to obey a state only when the justification for its authority follows from principles which form part of how we think about ourselves. When this fit obtains, states can enhance our ability to respond to reasons. We can be, therefore, more autonomous under political authorities even without giving them our actual consent.

There is another important lesson to be learned here, which is that we must be careful in the way that we use the term ‘legitimacy’. As I will discuss in Chapter 4, legitimacy has both a descriptive and a normative meaning. These must be carefully distinguished since the Weberian notion that political legitimacy consists solely in the positive attitudes of a population towards its leaders and institutions can be misused to lend support to the insufficiency thesis.[[37]](#footnote-37) I shall argue that for an authority to be legitimate with respect to an individual it must be possible for her to feel bound to obey that authority. It will thus be shown that there is a descriptive element involved in ascertaining the possibility of an individual feeling the binding force of obligation, but this is only part of a complete normative account.

In the next section I will argue that the sufficiency thesis helps itself to the characteristic binding force of obligation without due regard for the difficulties involved in explaining exactly what that entails and how it can be appropriately generated.

**1.6 – Autonomy, Anscombe, and the Binding Force of Obligation**

Perhaps all of this talk of voluntariness is missing the mark. After all, it is only autonomous agents who can act voluntarily. If autonomy can be appropriately respected in hypothetical consent theories without a standing commitment to voluntary action as the means by which political obligations are formed, then perhaps the insufficiency theorists may yet be placated.[[38]](#footnote-38) How might this be achieved? In one sense it seems straightforward. As we noted with Scanlon earlier, one of the considerations that factor into the justification of a coercive state is the ability of individuals to exercise their autonomy within their political community. But this does not mean that the general justification of political authority is subject to the actual agreement of anyone.

Instead, the value of autonomy is thought to be implicated in our relationship to sufficiently justified principles and this is cashed out in terms of the bindingness of those principles, the imperatival form that they take. There is a reason for this. It is that moral contract views such as those propounded by Rawls and Stark have a very specific conception of autonomy.[[39]](#footnote-39) To be autonomous is to be bound by the principles that express our nature as rational and reasonable beings. These are the principles that would have been selected by the representative parties in the original position. But now we have another mystery. What does it mean to say that we are bound to conform to some moral principle?[[40]](#footnote-40)

Permit me to sketch here Anscombe’s (1958) influential objection to a Kantian theory of moral obligation before we engage with her general critique of the concept of obligation in Chapter 2. Anscombe flatly rejects the idea that an idealised account of a rational self can stand in the appropriate relationship to the whole person to explain how that person is under an obligation to do or to refrain from doing something. This is because you cannot plausibly serve as your own commander. If I decide to abstain from alcohol for the month of January then even though I might insist to myself that I will follow through on my solemn vow it is always within my power to change my mind and let myself off the hook. How, then, can it be said that I am bound to stick to the rule that I have made for myself? There is no rational self, distinct from the lazy, selfish, sensible self. There is only the whole person, and she can surely not command herself.

There is a glimmer of hope for the contract theorist here because, on the Rawlsian scheme anyway, the representative parties are not special parts of ourselves. Rawls takes great pains to clarify that the representative parties are theoretical constructs.[[41]](#footnote-41) However it is, in the end, only a glimmer because although Anscombe’s particular objection to self-legislation does not directly apply, it is devastating nonetheless. If the binding force of obligation is not generated by the relationship between our two selves, i.e. nobody serves as commander, then where does it come from? In what sense do we *have* to conform to the decisions of the representative parties? Indeed, this turns out to be a general problem because the insufficiency theorist must eventually explain not only why, but how we are really bound by our voluntary actions.

In Chapters 2 and 3 I will consider this problem in more detail and develop a sanction theory of obligation that explains the binding force of obligation in terms of a very particular cost. This cost will turn out to be the unravelling of an individual’s conception of herself. On my account there is an important role for hypothetical consent in a social contract argument. It can be used to extend and expand a practical identity from certain core principles. If an agent is committed to the core principles underpinning the argument, and would indeed have consented to the more complete picture that the theory builds up, then the sanction of unintelligibility ties that agent to the conclusion of the argument as well. In response to Anscombe, I will argue that although an agent cannot serve as her own commander, she can police her commitments and exact a penalty in the event of a breach.

Voluntary agency, in the form of consent, is relevant to some of our obligations because it is important to us to have the normative power to assume various kinds responsibilities in our personal relationships. It is relevant to others because we think that agents should have certain kinds of control over things that affect them. It is a mistake to think that consent is relevant to the generation of a duty of obedience to a suitably just state in either of these ways. There are compelling reasons for most of us to obey just states. On my view, however, our commitments to our practical identities can be voluntary and this is how the concern for voluntariness evident in the motivation of the insufficiency theorists and, indeed, the traditions of the social contract can be preserved. It is a further question why it matters that these commitments are voluntary. I will argue in Chapter 5 that it is because identifying with principles is an important method of responding to reasons that it matters that we voluntarily bind ourselves to our political obligations.

**1.7 – Conclusion**  
  
For an insufficiency theorist, the hypothetical consent model of the social contract fails to generate political obligations. A contract theorist can respond along the lines that Stark suggests, but if she does then she faces three serious problems. First, she must reconsider and possibly even excise the traditional deference to an ideal of voluntariness that characterises the social contract tradition. This is easier said than done, and even if it can be achieved it will make a contract view less attractive as a result. Second, she must explain away the conviction that voluntary agency really does have something important to do with our political obligations. This leads to the third problem, which is that if she attempts to sate our voluntarist intuitions by appealing to autonomy she exposes herself to a deep worry about obligation in general, namely; in what sense are we bound to conform to moral principles? This, I have argued, is particularly pressing if she relies on a view of autonomy that explains our relationship to moral principles in terms of our hypothetical consent. Obligation, it seems, squirms back out from under the thumb of the contract theorist.

We began with a niggly little worry about the source of political obligation in contract theories that employ hypothetical consent. However, we have ended up with a much deeper general concern about the binding force of moral obligation. In this dissertation, I will address both problems and show how they are connected by a need for a clearer conception of the role of hypothetical consent.

I will develop a sanction theory of obligation that explains the binding force of obligation through the cost of unravelling one’s practical identity. To establish somebody’s hypothetical consent to a moral principle is, on my view, to invite her to exact a cost from herself in the event that she breaches that principle. Our commitment to our practical identities, I will argue, can be a voluntary one. In this way, a contract theorist can reconcile voluntarist concerns about the relationship between the individual and her obligations with the use of hypothetical, rather than actual, consent as the standard of their justification.

Formulating a new theory of obligation will not work in isolation, however. In particular, it raises important questions about how we should understand legitimacy and autonomy. I will, therefore, expound theories of legitimacy and autonomy that incorporate and expand on the insights garnered from my discussion of obligation. The final chapter of this dissertation will then return to the social contract and argue that it must be inverted in order to focus on the identity that a prospective citizen must have if she is to relate to the directives of her state in an obligatory fashion.

# 2

# *The Cost of Breaching an Obligation*

**2.1 – Introduction**

Oscar Wilde once quipped that duty is what we expect of others.[[42]](#footnote-42) But do we expect any less of ourselves? In this chapter I am going to argue that it is in the demands that we place on ourselves that we can locate the distinctive phenomenology of being bound that we associate with the concept of obligation.[[43]](#footnote-43) My claim will be that feeling the force of an obligation to do or refrain from doing something consists in requiring our own selves to do or refrain from doing that thing, on pain of unravelling our self-conceptions as reason-responsive agents committed to a hierarchy of exclusionary principles. This self-understanding matters to those of us who want to live what I shall call an intelligible life. The binding force of duty will thus be rooted in the cost that follows from any breach of those principles to which an individual is committed. This sanction may not be noticed or felt by an individual, but I will argue that what matters is that she weakens her grip on her own commitments and, ultimately, her self-image and it is against this that the call of duty first whispers, nags, pricks, and prods.

This move is prompted by the difficulty of establishing obligation through a social contract that depends on hypothetical consent. How is it that we are *required* to do certain things and refrain from others on the basis of a contract that nobody signed? The broadly Kantian answer, that establishing such hypothetical consent is to establish that our rational selves have legislated for us in a manner consistent with the store we set in autonomy, will not do. As Anscombe (1958) points out over and over again, this notion of self-legislation is absurd. The fiction of a distinct, rational self cannot stand in the necessary relationship to the whole person. However, this does not exhaust the possibilities open to the contract theorist. While an individual cannot plausibly serve as his own commander in the sense of authoritatively setting the content of his moral obligations, he can conceivably be the entity who exacts the appropriate cost in the event of a breach.

I shall set aside here the task of fixing the content of our moral obligations and focus instead on the question of how the binding force of those commitments should be explicated on a contract view. A genuine moral duty must be sufficiently supported by reasons, but is only binding when an agent relates to it in the right way.[[44]](#footnote-44) My contention is that moral commitments are policed by individuals themselves. This, I will argue, moves hypothetical consent back to the centre of the stage because arguments that deploy hypothetical consent invite us to impose a sanction if we reject their conclusions. I will also show how this account allows us to restore the significance of voluntariness as the core of the social contract.

I will begin by recapping the challenge involved in deriving obligation from a broadly Rawlsian social contract argument, before discussing in Section 2.3 the conceptual terrain of moral obligation itself. Section 2.4 will then consider the nature of sanctions and the role of conscience, arguing that only a self-imposed

sanction will give us what we need. Section 2.5 goes on to defend the view that the appropriate cost of obligation should, therefore, be explained in terms of an individual’s undermining of her own self-conception and Section 2.6 elaborates on the notion of intelligibility that I employ here. I then turn, in Section 2.7, to address the ideal of voluntariness in more detail.

**2.2 – Social Contract Theory and Obligation**

There are two standard ways to get an obligation to obey a legitimate state directly out of a social contract theory. Either you ground it in the actual or in the hypothetical consent of the parties to the contract. In the first case, obligation is generated by a normative power that individuals are generally understood to possess. By giving consent each participant places herself under an obligation to abide by the terms of the contract. This obligation is, in the first place, owed to the other parties and they are entitled to demand satisfaction. Of course, there is very little reason to think that this model can serve as the basis for state legitimacy.[[45]](#footnote-45) If actual consent won’t get the job done then perhaps we can appeal instead to hypothetical consent, to the significance of establishing what people *would* have consented to under certain conditions.

However, hypothetical consent cannot ground obligations in the same way as actual consent. This is because hypothetical consent is, as we discussed in Section 1.4, not really consent at all. Regardless of whether it could be established that anybody would give their consent to some potential contract or agreement, Dworkin maintains that in the absence of their actual consent we simply have no permission to enforce such a contract.[[46]](#footnote-46) Hypothetical consent provides no argument beyond the merits of the contract itself so it cannot ground an obligation in the way that actual consent normally does.

So, we might just appeal to the arguments supporting the contract to which hypothetical parties are, hypothetically, giving their hypothetical consent. The force of these arguments would, on this picture, then have to be sufficient to ground obligations. But there is a worry about this strategy. Simmons argues that we should separate out two key elements of political philosophy. On the one hand we would have the task of deciding whether or not political authority is a good thing and what is the best variety to have. On the other, would be the challenge of establishing what sort of relationship or interaction would have to exist between an individual and the state in order to legitimise its imposition and enforcement of duties. Simmons (1999, p.752) thinks that these two issues should be kept apart for the sake of individual liberty: “To deny this is simply to deny the natural freedom of persons, a basic and plausible Lockean premise”.[[47]](#footnote-47)

Stark (2000) notes, however, that while we think that our actual consent is important to generate certain types of obligations this does not hold across the board.[[48]](#footnote-48) My agreement is indispensable if I am to be bound to proof-read my friend’s thesis, but not for my obligation to refrain from murdering people. Simmons has simply gotten mixed up. Our political obligations may seem to be up to us and thus something we must incur voluntarily, but this is not the case. For Stark, the standard of hypothetical consent is required for justifying moral principles; meeting this mark is just what it is for something to be a moral principle. As such, the use of hypothetical consent in their justification is not intended to mimic or replace the operation of actual consent. On this scheme, political obligations are derived from higher-order moral principles and this is why hypothetical consent can ground the authority of just states to impose and enforce laws.

So far, so good we might think. However, although Simmons does not have the objection that he thinks he has, it seems to me that his line of thought prompts two other interesting worries. First, can hypothetical consent really generate the bindingness for political obligations that typically accompanies more fundamental obligations? The backdoor approach to political obligations seems to leave something important behind. We feel the pull of the obligations that we create by giving our actual consent. We feel the pull of our obligations to refrain from theft and murder. But why don’t we feel the pull of our political obligations in the appropriate way? Why, indeed, are they so often portrayed as the kind of thing to which consent is relevant? On one level, this is an objection about the phenomenology and potency of our obligations. On another it has to do with the sense we can make of the concept of obligation itself. Responding to Dworkin’s indictment of hypothetical consent by interposing a view about the justification of moral principles opens the contract advocate up to criticism based on Anscombe’s disquiet about the coherency of moral obligation more generally.

For this reason, it would not help Stark to adopt an externalist stance on obligation or to try to explain away the relative impotence of political obligations with social or historical considerations.[[49]](#footnote-49) However, I do not think that either of these would be a good strategy for a contract theorist anyway. The appeal of the social contract depends not only on identifying the content of our obligations, but also on reconciling us to our commitments by putting them on a contractual footing. Rawls, in particular, is very keen to emphasise the idea that they should be seen as self-imposed. This, I think, is best understood as an attempt to secure their binding force in our deliberations.[[50]](#footnote-50) A key objective is to argue in such a way as to make us feel party to our own obligations and this will ideally include an account of how these obligations can gain some traction.

The second worry is related. The backdoor approach is not obviously consistent with the deference that is usually paid to the ideal of voluntariness. As we saw in Chapter 1, Stark (2000, p.334) is prepared to concede this point. However, I think that this is much too quick. There is more to be said here. Moreover, dropping the concern for voluntariness will have knock-on effects, particularly for how we understand legitimacy.[[51]](#footnote-51) This all suggests that we can do better if we demonstrate that tangible voluntary participation still has an important role to play in a complete theory of obligation.

I want to suggest a strategy that will secure this objective and at the same time resolve the question of the bindingness of moral duties established by hypothetical consent. Let’s say that the content of our general moral obligations is fixed by hypothetical consent, as Stark suggests.[[52]](#footnote-52) This would then be sufficient to ground something like a justification right[[53]](#footnote-53) for just states to go about their business. However, obligations should ideally play a substantive role in our deliberations and they should *feel* binding. Joseph Raz (1986) understands duty in terms of what he calls pre-emptive or exclusionary reasons. These are reasons that cancel or block the force of other reasons, and to have one is, in a sense, to be thereby constrained. So, I will borrow this scheme, but I will add the idea of a sanction in order to explain the bindingness of obligations – the unacceptability of breaching them.

It is up to each individual to impose this sanction on herself, insofar as she endorses the exclusionary aspect of the reason and cares about being the sort of person who responds to reasons in the appropriate way. This is what I mean by policing your own commitments. As I will explain in Section 2.6 below, the cost itself can be explained in terms of the damage that is inflicted on an individual’s self-conception as a reason-responsive agent living an intelligible life. If you fail to meet your commitments then you are not the person that you thought you were. This is a distinctly uncomfortable position to be in and that provides the core of the sanction. There are additional disagreeable consequences to undermining one’s own self-understanding. For instance, it becomes more difficult to make promises and set resolutions in the future. However, what matters for my purposes here is the original, abstract penalty that follows from unpicking one’s self-conception.

The natural objection here is that a sanction theory of the force of obligation relies on inappropriate motivations since it speaks to an individual’s self-interest. I will deal with this objection in Chapter 3. For now I will focus on outlining my view and on explaining two important points in its favour. First, it can meet Anscombe’s challenge and substantiate the characteristic bindingness of obligation. Second, since individuals play the key role in supervising their own commitments, we can pry open the conceptual space to reintroduce a meaningful concern for voluntariness that, recall, would otherwise be hollowed out in a contract methodology. But before I discuss this proposal in more detail I will take a step back in order to frame the general topic of moral obligation.

**2.3 – Desiderata of a Theory of Moral Obligation**

A complete theory of moral obligation would be an ambitious project. Before I carve out the precise territory I want to cover here, it will be useful to sketch what it is that we ideally want from a full theory. By what criteria could we judge success or failure in that endeavour and what considerations must be borne in mind as we tackle more particular issues? The goal, of course, would be to establish exactly what obligations are, what kinds of creatures they apply to, and what it is to have one.

More specifically, important tasks would include describing the motivational role of our obligations, vindicating that role, explaining how duties can be generated and discharged, and establishing our epistemological access to our obligations. We would also need to find a principled way to distinguish moral obligations, which are often thought to be importantly different to other senses of ‘ought’ that depend on our holding a particular interest, such as proper functioning oughts (you ought to water that flower),[[54]](#footnote-54) conventional oughts (you ought to use this fork for salad and that one for dessert), and prudential oughts (you ought to invest your money rather than leaving it in the bank). How do these different kinds of obligation relate to each other and is it worth our while trying to group them all under a single concept?[[55]](#footnote-55)

A full theory would also deal with cases where obligations appear to come into conflict. For instance, is W. D. Ross (1930) right about the relationship between *prima facie* and all-things-considered duties?[[56]](#footnote-56) It would determine the possibility of obligations to self[[57]](#footnote-57) and distinguish between those obligations that we acquire and others that apply regardless of the things that we do in our lives. It would consider the significance of the various roles we occupy over the course of our private and public lives and establish what licenses the imposition of external sanctions such as blame and punishment.[[58]](#footnote-58) Finally, it would account for the bindingness of obligation and capture that distinctive sense of inescapability that typifies the acknowledgment of a duty. It is this last topic with which I will be primarily concerned here.

There are two additional, and related, concerns that I think we should also try to accommodate, and which I will address in the next chapter. They stem from the decidedly odd phenomenology of obligation. Under an obligation one feels compelled. One feels restrained, bound, unfree. It would be natural to expect that we would view such requirements as a burden, perhaps a necessary one, to be suffered and endured for the sake of various personal and social goods. However, rather than always wishing to be free of our obligations, in our nobler moments we sometimes aspire to be the kind of person who meets her commitments cheerfully. To endorse our obligations and assimilate them into our motivational set is to become, in a sense, a better person, and most of us want to be better people. The question then arises: if we discharged our obligations because we wanted to, without any reference to their binding force, would we really have obligations at all? However, this would be an odd conclusion indeed. So, it would be better if we could explain how our obligations can retain their force, even when we’ve moved beyond moral continence to a state approaching what Schiller described as grace.[[59]](#footnote-59)

A very similar thought is that even when we feel the weight of duty pressing down, compliance is not exactly dragged out of us. Feeling the binding force of duty is not incompatible with doing something for the ‘right’ reason. Feeling an obligation often takes the form of one’s attention being repeatedly directed to the decisive reasons involved.[[60]](#footnote-60) There may be a sanction hovering over us if we fail to discharge the obligation, but ideally this should be compatible with acting on the basis of those reasons. As I suggested earlier, it would seem to get our relationship with the salient reasons in moral scenarios all wrong if we were moved purely by the fear of some unacceptable consequence for ourselves.

So, a complete theory would not be an easy ask. However, bewildering as the conceptual terrain may perhaps seem, there are, I think, three key questions that the debate can be structured around:

1. Where do we get the content of our obligations?
2. What licenses punishment and/or blame when we fail to do our duty?[[61]](#footnote-61)
3. In what sense are our obligations binding?

My concern here is primarily with the third question. A satisfactory theory of obligation must acknowledge that we do not simply have moral commitments, we *relate* to our reasons in various ways and to be under a duty is perhaps the most important example of this.[[62]](#footnote-62) Before I broach this subject in more detail, there is an important challenge to the intelligibility of the concept of obligation itself in modern moral philosophy and it is to this challenge that I must first turn.

**2.4 – Command and Commanders**

Alasdair MacIntyre powerfully expresses Anscombe’s chief worry about the concept of moral obligation via an imaginary dystopian future.[[63]](#footnote-63) He asks us to picture the aftermath of a philistine political movement. The ‘Know-Nothing’ brigade made a serious job of erasing our scientific knowledge and abolishing the spirit of enquiry itself. Although some enlightened souls eventually fight back and try to piece together the remaining fragments, the enterprise was always doomed to failure. Too much was lost. In particular, the foundational beliefs that the natural sciences needed in order to hang together as a unified, intelligible scheme. MacIntyre (2007, p.2) thinks that this scenario is analogous to our relationship to central parts of our moral discourse: “What we possess, if this view is true, are fragments of a conceptual scheme, parts which now lack those contexts from which their significance derived. We possess indeed simulacra of morality, we continue to use many of the key expressions. But we have – very largely, if not entirely – lost our comprehension, both theoretical and practical of morality”.

MacIntyre reaches back to the Greeks for a teleological foundation to ground our moral discourse. In this he is different to Anscombe, who locates the modern malaise in the decline of the Judeo-Christian worldview. On her diagnosis, moral philosophy continues to operate within the paradigm of law but this idea has been cut loose from its conceptual tether.[[64]](#footnote-64) The worry is that it is a concept rooted in a religious worldview in which God lays down the law and we follow it. He commands and we obey. Take God out of the picture and it is a bit like taking the battery out of the bunny, it stops working.[[65]](#footnote-65) Two alternative strategies have emerged: have something else play the role of commander or else revise the concept so that it doesn’t need one.[[66]](#footnote-66) In this section I will consider so-called ‘no-command’ theories of obligation[[67]](#footnote-67) in order to bring out the case for and against building a commander into a theory of obligation. I will then be in a position to argue that, so far as bindingness is concerned, the debate is misguided, the issue really revolves around the extraction of the cost of obligation rather than the authority to impose and enforce duties.

On a no-command view, we can be said to have an obligation when there are sufficiently weighty moral reasons for performing (or refraining from) some action. On the strength of these reasons we are licensed to say that the action is either required or impermissible. And at first glance this looks like a plausible view. Unless something has gone pretty horribly wrong we don’t need a commander to tell us that murder is wrong or to bully us into returning the wallet we find on the street to its rightful owner. Most of us, most of the time, think that something’s being the right thing to do is sufficient reason to do it.

Tempting though it might seem to leave it at that, the no-command view faces an obvious challenge. In the absence of someone to whom we owe our compliance with our moral reasons, what sense can we make of this idea that we are *required* to do something? I have some very good reasons to shift my diet in a more vegetable-friendly direction, but this hardly constitutes a requirement. Perhaps moral reasons are different. Susan Wolf (2009, pp.351-352) considers the possibility that moral reasons not only count, but count decisively to create obligations: “That is, they outweigh all other reasons that might favour doing anything else. To say one is morally required to do something in this sense, would imply that, all things considered, you should do it, in a situation in which the salient reasons for this judgment are ones of a moral sort”. But, as she suggests, this continues to dodge the question. What is it about the weight, even the overwhelming weight, of some reasons that can transform a superior or optimal course of action into a mandatory one?

One way of putting some flesh on the bones of a no-command theory is to apply Raz’s theory of exclusionary reasons. For Raz, these reasons operate at one remove from our first-order reasons for doing, or refraining from doing, something. They replace the first-order ones by imposing a more efficient structure on our reasoning and should not be understood simply as contributing more weighty considerations to the balance of reasons.

A soldier has an obligation, for Raz, to obey his commanding officer because it is no longer up to the soldier to decide whether or not he’d rather curl up with a good book by the fire than get down in the mud to do another twenty press-ups. The satisfaction of reading the book or the pleasant warmth of the fire are first-order factors that would ordinarily count against strenuous exercise in the cold and the rain, but they are rendered inert by the structure that rationality imposes on his deliberations because of the greater strength of the reasons that count not only in favour of the soldier doing the twenty press-ups, but also in favour of the officer’s being in a position to demand that the soldier do them.[[68]](#footnote-68) This is not to say that the officer can order the soldier to do anything at all. There are important conditions of legitimacy on any authority since authorities must serve the interests, broadly construed, of those over whom they claim jurisdiction. This is why he calls his account of authority the ‘service conception’.[[69]](#footnote-69) The officer in our example would not be helping the soldier to better comply with his reasons by ordering him to torture prisoners, say, so it follows that the officer would be acting *ultra vires* and the relationship would have to be opened up to scrutiny on the first-order level.[[70]](#footnote-70)

They key for Raz is that when things are running smoothly the soldier better complies with his reasons overall by trading in his first-order reasons for intermediate ones and acting on foot of those. When his commanders issue orders they are giving him intermediate reasons that exclude some first-order considerations. Obligations in general are then supposed to take this shape, even if they are not owed to a single identifiable individual such as a commanding officer. The duty to respect others’ bodily integrity, for instance, just is an intermediary reason that excludes, among many other, more significant things, the pleasure I would take in flicking my brother’s ear when he is not paying attention. Rather than comparing strength, conflicts between levels are avoided “by a general principle of practical reasoning which determines that exclusionary reasons always prevail” (Raz 1990, p.40). So, failing to discharge our obligations is the same thing as failing to meet an important standard for rational agency. But why should this matter?[[71]](#footnote-71) The question is this: what binds us to this particular way of being a rational agent? Raz does not entertain this question but it is a significant one if we are concerned with the special phenomenology of bindingness.

Here is another problem with the Razian position as it stands. However sensible it would be as a deliberative strategy to look to my intermediary reasons and to disregard the considerations that are ruled out, this does not prevent me from being led astray by attending selectively to those considerations that have been ostensibly excluded. Let’s say that I would do better to exclude the deliciousness of cake as a factor when I decide what to eat for lunch. Raz is quite right that we require a way to transform the structure of our reasoning so that my exclusionary reason to not eat cake is privileged over those features of the cake that would otherwise count in favour of eating it. But as anyone who has ever thought that they shouldn’t eat a piece of cake, then eaten the piece of cake, and then regretted it afterwards can tell you; a general principle of practical reasoning is precisely no help at all when it comes to resisting delicious, cake-based temptation. Selectively attending to some considerations rather than others is something that human beings have become very good at.

This might seem to miss the point. After all, Raz is talking about the nature of reasons and I have countered with a human psychological weakness. What makes the point relevant is that we are interested here in understanding what it means to be bound to discharge an obligation. What Raz gives us is a more efficient scheme for organising our reasons. And I think that he is correct to think that a very large piece of the obligation puzzle has to do with how first-order reasons relate to higher-order reasons. What he does not give us, however, is a plausible story about how we are compelled to act, or to refrain from acting, in specific ways.

This is where we might think that it helps to introduce a commander. However the content of our obligations is determined, there could be a role for a commander to make what we have sufficient reason to do obligatory by wielding rightful authority. Being required consists, on this picture, in owing performance to someone in particular who is entitled to make demands. The trouble with this, though, is that it is extremely difficult to see why the emphasis is placed on the presence of a commander rather than the power that a commander can bring to bear. It seems more natural to focus on the basic idea that to be under an obligation is to have your freedom limited in some important respect, but if that is right then the question of bindingness actually turns on the barriers or costs that we might understand as limiting or constraining our freedom.

Perhaps this is too quick. There are two possible ways in which a commander might explain the special character of obligation. One is that the commander could have the power to compel compliance. This approach quickly collapses into a sanction view since it is the commander’s ability use force of one kind or another that explains the inescapability of the obligation. The other option is to locate the requirement in the relationship between the individual under the obligation and the commander to whom she owes that obligation. On this account it is the nature of the relationship that explains the constraint. However, this option falls prey to the same objections as the no-command view. It may well be the case that a person has very good reason to do what she is told, but the strength of those reasons does not transform the structure of her deliberations so that the intermediary reason to obey the commander is tangibly privileged. That still requires a sanction, and that is what I shall discuss next.

**2.5 – Sanctions and Conscience**

In her demolition job on the concept of obligation, Anscombe (1958, p.2) dispatched the idea of conscience in exactly one sentence: “Butler exalts conscience, but appears ignorant that a man’s conscience may tell him to do the vilest things”.[[72]](#footnote-72) So far as this goes it is, of course, true. All manner of unpleasantness has been sanctioned by the little voice inside the head. However, we do not have to cast the role of conscience as commanding or identifying our moral obligations. Rather, I submit, it is to play the part of enforcer. That is why conscience is often depicted as a burden, an external inconvenience. It is not a driving force. Rather, it is an anchor and like any heavy object it runs the risk of becoming a millstone if it is not properly deployed.

Mill (1979, pp.303-304) pointed out that we “do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience… Duty is a thing which may be *exacted* from a person as one exacts a debt”. The foundation of a sanction theory of obligation is this notion that wrongdoers can have no complaint if they are made to bear a cost for their transgressions. Of course, sometimes punishment may be counter-productive and sometimes it may be eluded. The point, then, is not that some sanction necessarily does follow, but that some sanction is appropriate.

Now, Mill starts with law and works his way down to conscience but this is misleading because it should really be the other way around. The most basic sanction is the one that individuals exact from themselves in the solitude of their own thoughts. In the first place, it is basic because it is ubiquitous. Even if more public or more severe sanctions are justified, we also expect that the wrongdoer should reproach himself. And if he does not mete out some internal punishment then there is a worry about whether he is really bound in the right way. This is why demonstration of genuine remorse is so important in criminal justice. It exhibits sensitivity to the pain or suffering that has been inflicted, but it also serves as a sign that the individual’s future behaviour is checked by something deeper than physical or social constraints. Even if he were assured that these would be avoided, his freedom would still be curtailed because the constraint has become a part of his own psychology.

It may seem that what I’m after here are the so-called moral emotions of, say, guilt or shame. However, these are merely additional, more concrete forms that a sanction may take. They are not fundamental because I can sometimes violate my obligations without it being fitting that I feel guilty or ashamed. How am I bound to stop at a red light on a deserted road at three in the morning?[[73]](#footnote-73) It is not because I should feel guilty for a harmless violation of the traffic laws.[[74]](#footnote-74) Nor would it be appropriate to be ashamed of myself for not twiddling my thumbs until the light changed. Underlying the anguish of the moral emotions must be a more abstract cost.

The way that exclusionary reasons operate is helpful here because they impose a structure on our first-order reasons. At any given moment there are all kinds of considerations that count in favour of doing all sorts of things. Right now I’m trying to organise some ideas and turn them into something approaching a chapter of a philosophy dissertation but, of course, there are any number of other things that I could be doing. I could, for instance, put on the tennis, or pick up my guitar for a quick strum, or head out to the garden to drink a nice cold beer, or read my book, or take off to the Peaks, or stare out the window, or cut my nails, and so on and so on. Instead I keep tapping away on the keyboard. I have committed to doing some work today, and that means that the enjoyment or distraction that I would get from doing these other things should not count in their favour, not today.

For Raz, it is simply in the nature of second-order reasons to subsume and replace first order reasons. They bind on account of their scope. We do better by our first-order reasons by attending to the second-order ones and that is why we should adhere to them. But what happens when we fail in this regard? I think that what keeps me sitting at my desk is not simply the rationality of my second-order reason to ignore the lure of temptation. Rather, it is the deep desire to be the sort of person who sticks to his second-order reasons. If I allowed myself to pursue every shiny distraction that came across my path I would, in Harry Frankfurt’s (1971) terms, be a wanton, and I don’t want to be a wanton. Why not? Well, a wanton is not the kind of being who can live a life that has anything more than fleeting moments of meaning. Only by imposing a structure on our first-order reasons can we build up a life that is intelligible as a whole and every time we do something that wobbles that edifice we make it more difficult to make sense of ourselves.

It is a little bit of hyperbole to pose the choice as one between being a wanton and being a maximally resolute character, firmly attached to his higher-order reasons. It is possible to be a bit feckless while maintaining an identifiable character. But it is not possible to be feckless in all regards and at the same time to respond to reasons by committing to exclusionary principles for action.[[75]](#footnote-75) Indeed, fecklessness is worrying from this perspective because the effectiveness of exclusionary reasons relies on a disposition to stick to them.

This will all sound more than a little mysterious, but I think that it can be illustrated quite straightforwardly. If I make a solemn promise to pick my friend up from the airport and I succumb to the temptation to play video games instead then I am not the kind of person that I thought I was. I find this realisation unsettling in itself and want to resolve it. The problem is that I have made myself unintelligible.[[76]](#footnote-76) This is the core of the sanction that I am proposing.

It is also likely to spawn more familiar problems of deficient agency. Not only will my friend have trouble trusting me again, I will have trouble making a serious commitment because I now have good cause to question whether I will stick to it. Similarly, if I decide to give up drinking for a month and fall off the proverbial wagon the next day, then I have raised a pretty serious doubt about my ability to follow through on my commitments.

But, we might well object, is this the right sort of cost to underpin moral obligation? Wolf (2009, pp.352-353) argues that mere failures of rationality are not enough to support either the external sanctions of punishment and blame or the internal agony of guilt and shame. My response is that the nature and severity of additional sanctions depend on the nature of the offence and are justified in that context. All that I am trying to establish here is the most basic sanction that is appropriate in all cases. It is also worth noting that it is not exactly a failure of rationality, it is more a failure of will or character. Finally, this sort of cost does not jar with the idea that we should meet our obligations because they are the right thing to do. I will argue in the next chapter that since how we conceive of ourselves is bound up with the principles that we recognise and the exclusionary precedence that they take over our prospective reasons, the binding force of my sanction theory actually harmonises with the reasons for being so bound in the first place. This explains the attention-directed element of my proposal. I return again and again to the considerations that support an obligation.

What all of this means is that *we tie ourselves* to our moral obligations by endorsing them as exclusionary reasons and staking our self-understanding on treating them as such.[[77]](#footnote-77) What commitments can be best justified is a related, but distinct element of a complete theory of obligation. This is the best sense I can give to the thought that we are, fundamentally, bound by our obligations. Hypothetical consent is not, as we established earlier, actual consent. But my hypothetical consent to some principle could be enough to make me feel that I ought to endorse it. I would have agreed, presumably, on account of the arrangement of my various beliefs and values. Once this becomes apparent to me, and on the assumption that nothing happens in the meantime to change the situation, then it would be very odd indeed to withhold my actual assent. More importantly, if the strength of my commitments is sufficient to generate my hypothetical consent then in the right circumstances it should be enough to generate my actual consent. If it did not then I would have to have a different relationship to my own commitments. Failing to give that consent would, therefore, undermine my self-understanding in exactly the way I have suggested so there is a cost, or sanction, attached to rejecting the salience of what I would have done. My actual commitments can thus bind me to my hypothetical consent. This is not, I think, an insignificant result. It will help to elucidate it by exploring the idea of intelligibility involved here in the next section before I tackle the tricky question of voluntariness.

**2.6 – Intelligibility**

Can we be more precise about this notion of the intelligibility of a person to herself and why that should matter? I think that we can. For one thing it is important to be clear that it is possible to make sense of an entity very like us but who acts from moment to moment.[[78]](#footnote-78) Such an agent could still act for reasons, in Scanlon’s sense that there are considerations it views as counting in favour of one action or another.[[79]](#footnote-79) Insofar as this is the case, those actions can be explained in terms of reasons[[80]](#footnote-80) and evaluated in that context. Indeed, some of the decisions that we make would seem to take this form. It does not, for example, take a complicated story about diachronic agency to work out my decision to shield my eyes from the bright sun.

What does require a more complicated story is the significance most of us attach to responding to reasons in a much more comprehensive manner. On the assumption that we really do persist over time,[[81]](#footnote-81) it is possible for us to live a life organised around an ideal that Lenman (2009) calls “the ideal of normative unification”. One of the chief attractions of living a unified life is that it opens up another way of responding to reasons. We can respond to them by directing our lives in various ways. Outwardly this may involve pursuing a particular degree at university in order to acquire the knowledge and qualifications required to become, say, an engineer or engaging in intense martial arts training if the goal is to win the Ultimate Fighting Championship. Inwardly, which is more interesting for our purposes here, it will involve deciding how best to think of oneself.[[82]](#footnote-82)

Your conception of yourself is formed partly by gaining an understanding of your tastes, preferences, and abilities; partly by ascertaining what is expected of you by those around you; partly by determining the values according to which you think people should live; and, finally, by arranging them all together to form something approximating a coherent whole. We are, most of us at any rate, engaged in the project of living a good and decent life. Just as it would be foolish for NASA to send an astronaut into space without a clearly-defined set of operating procedures for dealing with the kinds of challenges and problems likely to befall your average space traveller, so too would it be unwise for a person to head out the front door without developing her own set of operating procedures for navigating in the world.

The difference is that my operating procedures are not stored in a manual that can be consulted when needed. Rather, they are incorporated into my identity. We sometimes describe our commitments by saying that we ‘have a rule’ about something but this is an awkward way of getting at an elegant solution. It is not, strictly speaking, the case that I have a rule about helping little old ladies cross the road. Instead I simply understand myself as the kind of person who helps little old ladies across the road. A principle such as this is exclusionary in the way that Raz suggests. That I could shave a few seconds off my travelling time by ignoring my fellow pedestrian will not count as a reason against assisting her. There will, it is important to stress, be some considerations that are not excluded, but this is to be expected.

Were a person to violate a principle to which she was committed as a part of her identity, she would face the following consequence: she would not be who she thought she was. An individual is unintelligible to herself on my account when her actions are at odds with the principles that together constitute her self-conception. This kind of tension is something that almost all human beings are, as a matter of empirical fact, dead set against and that is why it can provide the tangible force required to get a sanction theory of obligation off the ground.

One might object that this view makes it impossible for people’s identities to grow or change, that it appears as if people should stick to a principle once adopted, no matter how misguided, on pain of unintelligibility. This is not the case. It is within our power to revise and extend our identities in light of better reasoning and lived experience. However, such change takes place within the context set by all of our other commitments. It is never a live option to ditch one’s entire identity and start again. Change on this level is a gradual and piecemeal process. It is also worth noting that many people do find it painful and difficult to alter their identities in response to new evidence or shifting social norms and practices. Making and remaking the self is a profound undertaking.

But how do we come by our commitments in the first place? Many of them are impressed upon us by our families and wider society. Can an individual’s identity ever be voluntary in any meaningful or important sense? And what sense of voluntariness is relevant to my development of contract theory? I shall examine this subject now.

**2.7 – Voluntariness**

I have said that a concern for voluntariness is an important part of the social contract tradition, but why should this be so? Why does voluntariness matter? Indeed, as I outlined in the previous chapter, on the view that Stark takes about the moral foundations of a contract procedure, a general concern for voluntariness is irrelevant to the operation of hypothetical consent and the justification of moral principles.[[83]](#footnote-83) However, the key to appreciating the potential of a reimagined contract approach is understanding how an individual can be reconciled to principles that it would make sense for her to endorse. It matters, then, that she commits to those principles of her own accord.

This is for two reasons. First, it is only by incorporating principles into one’s own self-conception that one can be tangibly bound to them in the appropriate way, as discussed above. And, second, we become more autonomous as more of our commitments are voluntarily selected or retrospectively endorsed. In this way we can gradually construct an intelligible identity by responding to reasons in a consistent and ordered fashion. I will have much more to say about this in Chapters 4 and 5 where I discuss legitimacy and autonomy in detail. In this section I will explain and adapt Serena Olsaretti’s account of voluntariness to show how our commitment to our practical identities can be voluntary.[[84]](#footnote-84)

Since nobody ever actually consents to the principles established via hypothetical consent, what would support the judgment that a role for voluntary participation is somehow preserved? Olsaretti argues that a choice is voluntary “if and only if it is not made *because* there is no acceptable alternative to it” (2004, p.139, original emphasis). The core of Olsaretti’s account is the idea that we must look to an agent’s motivations to understand the nature of her relationship to her choices: “it is agents’ attitudes towards the options they face – rather than the objective nature of those options, independently of the agents’ attitudes – that determine whether or not they act voluntarily” (2004, p.154).[[85]](#footnote-85) I think this is right. Since on my view, it is individuals themselves who impose the sanction, the idea is that the bindingness of a moral obligation is voluntarily enforced if it is not imposed *because* there is no acceptable alternative to it.

There is an important wrinkle here that needs to be ironed out. As Ben Colburn (2008) points out, it is critical to the account that people know what their options are. If you are not aware that you have an acceptable option along with your unacceptable ones then your choice cannot be a voluntary one. Olsaretti welcomes this thought and amends her definition as follows: “a choice is non-voluntary if and only if it is made because the alternatives *which the chooser believes she faces* are unacceptable” (2008, p.114, original emphasis).[[86]](#footnote-86) Whether or not an option is actually unacceptable is not, for either Olsaretti or Colburn, hostage to idiosyncratic values or expensive tastes. If, for example, a new breed of highwayman tries to hold me up by threatening to reveal key plot points of my favourite television show and thereby spoil it for me, he is not really forcing me to do anything. He does not succeed in coercing me because the option of keeping my valuables and having my enjoyment of the programme diminished is not sufficiently bad to count as unacceptable. I might think that it is, but I would be wrong. However, if a more traditional highwayman shoves a pistol in my face and tells me it’s my money or my life, my decision to hand over my valuables is forced, even if the pistol is not actually loaded. Even though we cannot ascribe voluntariness without referring to the perspective of the agent, that agent must share a defensible evaluation of her options as she is aware of them.[[87]](#footnote-87)

Colburn raises another objection that is instructive here. The objective standard that Olsaretti applies is an unhappily underdeveloped notion of well-being, which includes moral character – our lives go worse for us on her account if we perform morally bad actions. Colburn’s worry (2008, pp.103-106) is that moral actions come out as forced on this picture because it is unacceptable, for instance, not to save the drowning child.[[88]](#footnote-88) As such, he thinks that we could not be held responsible for our moral choices and it would be inappropriate to praise individuals for doing the right thing since doing the wrong thing would be so deplorable. His solution is to draw a distinction between prudential and moral unacceptability. The kind of unacceptability relevant for determining voluntariness is identified with the imminent potential for “serious, specific damage,” and it is the “prospect of awful personal consequences that acts as the instrument of force: it is because we cannot face such a loss of well-being that we have to choose in a particular way.” (Colburn 2008, p.105) He does not believe that the unacceptability of morally bad options generates palpable force in an analogous way.

In this chapter I have been trying to draw out the phenomenon of bindingness that is characteristic of moral obligation. Indeed, I have concluded that there is a very important sense in which we experience our duties precisely as non-voluntary and I have explained that in terms of a cost that an individual exacts from herself. Violating principles to which one is genuinely committed does raise the prospect of awful personal consequences and therefore is attended by tangible force. Nevertheless, Colburn picks up on an important tension that needs to be resolved between the phenomenology of obligation and the ascription of responsibility for moral actions.

If I am right that the binding force of obligation is constituted by a sanction, then we must accept that for a principle to manifest as an obligation it must be unacceptable not to honour it.[[89]](#footnote-89) What may be voluntary, though, is one’s original commitment to the principle under which the particular obligation falls. That principle can be endorsed because of the considerations that count in its favour; because of the reasons as they appear to you. And this explains something quite familiar about our responses to situations in which people do what they ought to do. It is more natural to compliment someone on what a good person they are after they do the right thing than it is to compliment them on doing a good thing.[[90]](#footnote-90) Perhaps we may do both, but it is much higher praise to laud someone’s character than to extol an act that they really ought to have performed anyway.

Consider, even if it seems a bit of a stretch, goalkeepers. The best goalkeepers perform the ostensibly straightforward tasks with a minimum of fuss. When a goalkeeper efficiently and competently makes a save that we would expect him to make we do not praise him directly for the save. Rather, we praise him for being solid and dependable, for being the kind of goalkeeper who does at least what is expected of him. Not making the save would be unacceptable, so in a sense it is no big deal that he made it. What is a big deal, however, is reaching a level of performance where we can put the proverbial house on him.

What ultimately matters is how a person relates to her commitments and whether they form part of her identity because of how she acknowledges the reasons that count in their favour.[[91]](#footnote-91) Insofar as she is persuaded that a principle is a good one then she positively embraces it and, crucially, understands herself as so constrained. As Olsaretti (2004, p.155) says: “so long as one option is acceptable and is one which the agent very much likes, the absence of an acceptable alternative may not be what motivates the agent”. I submit that it is possible for us to adopt principles for behaviour on the basis of their, for want of a better word, attractiveness, which follows from their being sufficiently justified.

So, to Colburn I can grant that to be bound by an obligation is to view a failure to discharge it as unacceptable. It really is despicable, for instance, to refuse to save the drowning child. At the same time I can maintain the possibility of a voluntary commitment to the principle under which it falls. I understand myself as the kind of person who would save a drowning child because I think that human beings are valuable and ought to be respected, and that this cashes out in saving them from drowning where possible.[[92]](#footnote-92) It is, therefore, perfectly appropriate to praise or condemn someone’s character if she is committed to it because that is the kind of person that she wants to be.[[93]](#footnote-93)

Now, most people want to live a certain kind of life, i.e. one that is intelligible as a unified whole, so how agreeable would it be to reject the pressure of your commitments? How acceptable, that is, would it be in the extreme case to become a wanton instead? Even if we don’t take it that far, how acceptable would it be have an inconsistent or unpredictable character? But of course this is the wrong way to look at it. For most people, it would indeed be unacceptable to become a wanton or to regularly eschew those principles in which they can see the sense, but they are not motivated by the repulsiveness of this possibility. I will argue in the next chapter that although there must be an appropriate cost if we are to say that individuals are really bound, the binding itself arises in the first place because of the persuasiveness of the reasons that support it.

**2.8 – Conclusion**

What is it to be bound by an obligation? My argument in this chapter has been that only a sanction theory can capture the element of constraint central to the concept of obligation. On my scheme, individuals who are interested in living a unified life police their own obligations and so a commander is not required to make the concept coherent. We understand ourselves as reason-responsive agents committed to a particular hierarchy of principles. Acting in ways that are incompatible with those principles has the undesirable consequence of weakening an individual’s grip on her identity. In effect, she makes her very self unintelligible. This provides the basis for a tangible sanction that can explain the unacceptability of violating a principle and, thereby, the bindingness of obligations.

Applying this insight to social contract theory opens up a new way to understand the force of hypothetical consent that is consistent with the deference paid to the ideal of voluntariness, an ideal that would otherwise be dropped. Insofar as we can construct (or retrospectively endorse) our identities it is critical to consider how we relate to them. Since this kind of commitment is not always undertaken because there is no acceptable alternative I conclude that it can be voluntary.

In the next chapter I will develop my theory of the binding force of obligation by reflecting on the significance of identity and address the most pressing objection to a sanction theory of the bindingness of obligation; that it appeals to self-interest and thereby weakens and corrupts our relationship to our moral reasons. I will then be in a position to explain how we can do the right thing for the right reasons even though faced by the prospect of a weighty sanction, and to defend the possibility of attaining a state of grace.

# 3

# *Sanctions & Self-Interest*

**3.1 – Introduction**

There is a deep suspicion that any theory of moral obligation that depends on sanctions to generate binding force ultimately reduces our most revered moral capacities to grubby self-interest. There is, I think, a very small grain of truth here, but it provides neither good reason to think that a sanction view must reduce morality to self-interest nor to exclude self-interest from our picture of morality in the first place. In this chapter I will show why reliance on a measure of self-interest does not pose a threat to an identity-based sanction theory.

In fact, there a number of distinct worries that cluster around the supposed threat of self-interest. I will deal with two here. The first is that a sanction view corrupts moral reasoning by introducing incentives to guide behaviour, while the second is that sanctions are too contingent to secure our moral commitments. This latter objection might be thought to be especially troubling for the view I defend since we have some control over the identities that we form and maintain. Perhaps we could simply change our commitments when they become a burden and so escape from under the sanction of unintelligibility. I will also consider another specific objection to the sanction of unintelligibility. We regularly make exceptions to rules that we set for ourselves without falling to pieces, so how weighty can this sanction really be?

With some assistance from Christine Korsgaard’s notion of a practical identity, I will develop the idea broached in the last chapter of an interest in living an intelligible life. The repugnant prospect of unspooling our identities generates a cost that most of us, most of the time, want to avoid. However, I will argue that the primary significance of this sanction is not that it provides a disincentive. Rather, it sharpens our focus on the question of who we want to be and, since we generally want to be the kinds of people who live intelligible lives, we are led back to the considerations that we judge to be salient in a given situation.

Changing one’s commitments is not something to be undertaken lightly. They are deeply embedded in one’s identity and to ignore them is often to precipitate intense self-scrutiny. On my view, one of the most important things that we are interested in is being the kind of person who acts for good reasons. We identify with particular principles as a response to reasons and we cannot successfully alter our identities without reference to, and support from, other elements of our self-conceptions. I will expand on this point to explain why it is also a mistake to think that making exceptions to your principles does not incur a cost and is nothing to worry about.

Responding to worries about rampant egoism also helps us to get a grip on two other important aspects of obligation that deserve more attention than they usually receive. I will defend a theory of the binding force of obligation that explains how it is that we can do the right thing for the right reasons even though we feel the weight of duty pressing down upon us. To feel the binding force of obligation is to have one’s attention drawn back time and again to the reasons that justify one’s allegiance to the relevant principle in the first place. I will also reflect on the possibility of grace and show how my theory permits a moral agent to embrace her commitments and do her duty cheerfully.

I shall start by outlining Korsgaard’s account of practical identity and identifying some important problems with it. Section 3.3 will then explain in more depth the objections clustered around self-interest, before I begin to respond to these challenges in Section 3.4 by expanding on the nature of the sanction that I propose and explicating how principles can acquire exclusionary force for individuals as part of a practical identity formed in the process of living an intelligible life. I will then be in a position, in Section 3.5, to evaluate the threat and show how a little self-interest is required to tie us to a hierarchy of regulative commitments that, together, can comprise a moral identity. When this is in place we can make sense of the idea of doing the right thing for the right reasons in spite of the pressure exerted by the cost of breaching an obligation. Finally, I will demonstrate that the sanction view I am recommending has the benefit of allowing for the possibility of attaining a state of grace.

**3.2 – Practical Identity**

In the last chapter I argued that the insistent nature of obligation, its binding force, must come from a sanction since it is only the prospect of bearing some cost that constitutes a genuine requirement. Further, the internal sanction of one’s own conscience must be the basic unit of currency in this debate since it is the only form that is ubiquitous when sanctions are appropriate.[[94]](#footnote-94) Moreover, because it is possible to violate some obligations in trivial cases without incurring or even justifying the unpleasant ‘moral’ emotions[[95]](#footnote-95) of shame and guilt, we need a more abstract cost, but one that is still weighty enough to underpin the potency of duty.[[96]](#footnote-96) In this vein, I suggested that to violate an obligation is to unpick your own self-image which is profoundly unsettling in its own right because it makes an individual unintelligible to herself,[[97]](#footnote-97) but also presents a practical problem insofar as this kind of failure makes it much more difficult to make and keep commitments. In this section I will motivate the threat of self-interest for a sanction-based theory of obligation by working through Korsgaard’s notion of practical identity. It will become clear that I disagree with her on some substantial points, but it is instructive to see both what she hopes to derive from the important ways in which we think of ourselves and why she ultimately denies that a sanction view on its own can explain the bindingness of obligations. So, what is a practical identity?

For Korsgaard, the deepest question we can ask concerns the source of normativity. Since we have the capacity to take a breath and reflect on our impulses, we are then faced with the problem of determining just how we *should* proceed. “The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go further” (Korsgaard 1996a, p.93, original emphasis). Since reflection generates the problem, the answer, she thinks, must also lie in reflection. If there are desires, say, that can survive the process of reflection then we will have found some reasons.[[98]](#footnote-98) But here’s the rub: how can a desire satisfy the scrutiny of reflection? Is there a reason to endorse it?

For assistance, Korsgaard turns to Kant. He comes at this problem from the perspective of free agency. A free agent is not determined by anything except her own will, but this is not to say that there are no constraints on what counts as a genuinely free action. There is exactly one: the requirement that you act on a principle, or maxim. Since nothing external to your will can determine what that principle must be, he concludes that what matters is the form of a principle, which is the form of a law. This is what leads him to the first formulation of the categorical imperative: the Formula of Universal Law.[[99]](#footnote-99) It is an extra step to get to the moral law since Kant’s moral law tells us to treat people as ends in themselves. This is a determinate principle and that is why it is possible to derive some recognisably moral content from it. The question, though, is how we can get from the first formulation to its later incarnations.[[100]](#footnote-100) The solution, Korsgaard tells us, is that the agent must think of herself as a citizen of the kingdom of ends.[[101]](#footnote-101)

To be self-conscious is, for Korsgaard, to be aware of yourself as something which is not reducible to your desires or impressions. “When you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desire to act on” (Korsgaard 1996a, p.100, original emphasis). To act on a principle under which, remember, desires can count as reasons is, therefore, to define this thing that is you. And just as we can act under different principles we can have different identities, each of which comes with its own set of reasons and, crucially for our purposes, commitments.

In all likelihood, dear reader, you are a human being. You probably have a family, a nationality, hobbies, a favourite sports team, an opinion on Blur vs. Oasis, and on and on. To be a Liverpool fan, for example, is to have a special relationship with the team such that you have a reason to cheer for them when they play, to boo Manchester United or Everton with almost equal vigour, and to engage in heated arguments about tactics and players at every opportunity. If you didn’t do these things then you wouldn’t be a fan, not really, and being a fan is one of the things that can give meaning to a person’s life. “The conception of one’s identity… is better understood as a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking” (Korsgaard 1996a, p.101). Similarly, Korsgaard thinks that to be a member of the kingdom of ends is to conceive of oneself as part of a community in which the members are entitled to expect respect and concern from one another, and to conceive of oneself in this way is to view oneself as having compelling reasons to act accordingly.

But even if we buy this story about what it is to have a reason as a bearer of a particular identity, how are we to understand the generation of obligation via this notion of practical identity? Rather than endorsement, Korsgaard (1996, p.102) points us to rejection, in the sense that we are obliged to do something when to do anything else would be abandon your identity and that, she tells us, “is to be for all practical purposes dead or worse than dead”. This is a pretty severe cost to bear, which is why we *must* reject actions that will prompt it.

Of course, some parts of our identities are deeper than others so I am obliged to save the drowning child, even if he is wearing a United jersey. Regulative priority is distinct from how much a particular identity may matter to a person. So, a hardcore fan might identify much more intensely with their responsibilities under that description but insofar as they also self-identify as a moral entity then that identity can play an administrative role.[[102]](#footnote-102)

However, Korsgaard wants a much stronger result than she can get from a sanction. Her ultimate intention is to ground a categorical morality and in order to guarantee this she moves away from the cost and pain of unravelling one’s identity and attempts to establish not only that a thinking self is in a position of authority over an acting self, but that this authority is the source of *moral* obligation. This means that Korgaard ultimately recommends a self-legislation view of moral obligation in which the rational self plays the role of commander even though she begins by talking up the cost of fracturing one’s identity: “It is the fact that we *command ourselves* to do what we find it would be a good idea to do” (1996a, pp.104-105, original emphasis).

Her basic argument is that since it matters to us to be able to justify our endorsement of any particular identity we must presuppose a bedrock identity as the kind of being who engages in reflective justification. Since this is prior to the reflective process, she argues that it is a condition of valuing any of the other ways of viewing ourselves that we value ourselves as rational creatures inhabiting a normative space. If we must value ourselves in that way then it is, she thinks, impossible to withhold that value from other rational agents who engage in reflective justification and so we get our foundational moral obligations.

Now, Korsgaard overreaches here in two important ways. First, she thinks that rational agents are compelled to form and respect maxims, but this is not at all obvious. We might object, as Nagel (1996, p.202) does, that: “If the will is self-determining, why can’t it determine itself in individual, disconnected choices, as well as according to some consistent law or system of reasons?”. His point is that we cannot ground categorical morality in a rational imperative to act under principles since it is perfectly possible for a free will to act in a momentary, fragmented way.

Second, she neglects to consider the possibility that a prospective reason may survive the test of rational scrutiny without terminating in the transcendental necessity of the value of rational agency. Sometimes we find that we care about something even if we can’t give a good reason for it. Indeed, sometimes we find that we care about things no matter how much we might wish that we didn’t, or how decisively the reasons stack up against them. Finally, we can choose to care about things and make our self-understanding dependent on following through on that value in everyday life. So, there may be things that I value simply because I care about them, and not because there is a special thinking self that is necessarily valuable and, therefore, has the power to set the whole chain of value in motion.

Responding directly to Korsgaard, G. A. Cohen (1996, p.175) makes a similar point, although he, like Bernard Williams, is concerned with having one thought too many.[[103]](#footnote-103) The objection here is not just that some values that form part of our practical identities will resist the challenge of reflective endorsement, it is rather that it would be inappropriate to value personal relationships, for example, because of a commitment to a principle that says we should value our personal relationships.

So, since this project is fraught with difficulties, what moves Korsgaard to supplement a sanction theory with a transcendental ground for a categorical morality? She thinks that an account of bindingness and a delineation of the appropriate subject matter is not enough to secure distinctively moral obligation. But why does she hold to this position? I have suggested that we leave questions of content to one side so that we can focus more precisely on the phenomenon of bindingness.[[104]](#footnote-104) That approach is not open to Korsgaard because on her picture the content of moral obligation must play a crucial role in generating its characteristic binding force.[[105]](#footnote-105)

Her major worry is that we cannot rely on a relationship that just happens to obtain between an individual and her commitments because she happens to conceive of herself as having a particular practical identity.[[106]](#footnote-106) The enemy is contingency: “Because these conceptions are contingent, one or another of them may be shed” (Korsgaard 1996a, p.120). An identity only generates binding force so long as the bearer continues to identify with it and self-conceptions can lose their grip on us in all sorts of ways. We can, for instance, grow out of some of our identities or pick one over another in cases where they come into conflict. This may be thought to be especially worrying since if loyalty to some self-conception comes at a high price then this very cost creates an incentive to find another way to conceive of oneself. Of course, if this is a live possibility then we shall have to devalue the currency of bindingness on the sanction view that I am offering, which would make it less attractive as a position.

Contingency is, however, just one of a cluster of objections that revolve around the role that self-interest plays in underpinning a sanction view. In addition, I will look at one other paradigmatic concern in this chapter: moral motivation. If a sanction is playing a substantive role in your moral deliberations then, precisely to that extent, you are motivated by the wrong thing because what you are worried about is the impact that it will have on you. What you should be worried about, of course, is figuring out the right thing to do. Worrying about what detrimental effect an action will have on your identity or self-understanding provides an inappropriate motive for an ostensibly moral action.[[107]](#footnote-107)

I will describe the threat of self-interest in the next section. Before I do, it is worth sketching the solution that I am going to offer. Korsgaard is right to suggest that the binding force of obligation comes from practical identities and the cost that follows for us when those identities are undermined or even shattered. Some actions have unacceptable consequences and that explains the sense in which we *must* reject them. However, she thinks that in order to get characteristically moral obligations you need to add some necessity to the mix so that, in order to be bound to do or to refrain from doing anything at all, you have to first value yourself as a special kind of rational agent and that is something that we cannot help but value. Now, rather than pursuing necessity in order to ground a categorical morality, I am going to stick with sanctions and focus on the unacceptability of a life in which we do not order our actions by forming and identifying with a hierarchy of principles. I will conclude that there is an interest that most of us have in living what I have called an intelligible life. This is something that matters profoundly to most of us, and it would be a very different mode of existence in which I didn’t care about being able to tell a story about why I’ve done the things that I’ve done.[[108]](#footnote-108) On my view, a will that determines itself in individual, disconnected choices is not necessarily an irrational one. It is, however, an unintelligible one.[[109]](#footnote-109) It is not a will about which you can tell a unifying story. It is not, in short, a will that has an identity. Even a will that has decided to live according to a principle of making disconnected choices has determined itself by that principle.

My account of the bindingness of obligations, therefore, does not capture everyone unconditionally and it relies on the importance that individuals themselves place on living a certain kind of life – thus is self-interest implicated. What may seem at first to be a source of weakness turns out to be a source of flexibility and strength. A small dose of self-interest emerges as the key to understanding how we can do the right thing for the right reasons even though the weight of duty presses down upon us, and it recedes into the background when we do our duty cheerfully, when we achieve a state of grace. But before I can demonstrate this we must first take a closer look at the threat that self-interest is thought to pose for a sanction-based theory of moral obligation.

**3.3 – The Threat of Self-Interest**

To begin with the potentially corrupting influence of sanctions. There is nothing generally wrong with a concern for the integrity of one’s practical identity. However, just as there is nothing generally wrong with trying to make a living, the possibility of a reward would be the wrong motivation for returning poor, lost Fido to his owners. In cases where one ought to act in some particular way we expect the moral status of the required or forbidden action to figure prominently in the agent’s thinking. Ideally, moral agents are moved by the reasons that count in favour of establishing a duty in the first place. The addition of a sanction may well provide a more reliable spur to action, but we might think that it impoverishes the moral character of that action.

Making this point, Nagel (1996, p.206) says: “To decide from the reflective standpoint what to do you have eventually to stop thinking about yourself and think instead about the question at issue… The answers to such questions may partly determine your identity, but they don’t derive from it”.[[110]](#footnote-110) And there is indeed a genuine concern here. If someone is always concerned only about ‘what’s in it for him,’ then he appears to disregard the considerations that are actually salient. But Nagel moves too quickly here. For a start, there are two ways in which we can understand his objection.

We might think that he is entreating us to engage in moral deliberation without entertaining any thoughts that reference ourselves or our interests. But this is obviously too demanding, partly because demandingness may be something that we are entitled to consider, but also because I can think of nothing wrong with appealing to my own self-image to stiffen my resolve if I can tell that it is waning. Jeanette Kennett (1993) argues that we can mix our motivations while still acting on the ‘right’ values. Indeed, commitment “to a value brings with it a willingness to work on our motivation and make use of the motivational resources available to us” (Kennett 1993, pp.264-265). There is nothing wrong with asking if this is the person that I want to be or if I could live with myself if I find that I am tempted to do something that I think is wrong, and it would be remiss of us to neglect such an effective strategy.

However, Nagel is really articulating his own theory of normativity and he thinks that we are compelled to adopt an impersonal standpoint once we realise that we matter only if other people matter.[[111]](#footnote-111) In order to determine what we should do, we need to abstract away from our individual perspectives and survey the scene in a less subjective way. In this way we can set aside the urgency of our personal interests and put them in a broader context, attending to the other values that most of us take to be morally salient. A sanction view might compromise that project if you think that it is designed to work by latching on to the urgency of private concerns and circumventing impartiality in order to channel self-interest towards ostensibly good ends.

Returning to Korsgaard’s concern with the contingency of sanctions in the absence of an unavoidable identity, a related problem that critics raise is the possibility that if you could only find a way to evade the sanctions then it would turn out that you were never really bound by them at all. This is particularly troubling for identity-based views that rely on individuals to impose internal sanctions on themselves. Cohen (1996), for example, worries that there is a sense in which you are free to cast off your obligation because if you make the law then you can unmake it and thus escape from under it.[[112]](#footnote-112)

And, as if this wasn’t enough, we need to face the fact that we make exceptions for ourselves all the time and yet maintain fairly secure identities. It would seem that a little transgression isn’t so bad; the cost of unravelling one’s self-image doesn’t seem to apply in lots of instances and it certainly isn’t prohibitive in others – otherwise we would have a psychological epidemic on our hands. In order to respond to these challenges we need to think about how a sanction view would operate, and what it is that we expect from an account of the binding force of obligation. The key is to understand how we can face a tangible sanction and yet act for the ‘right’ reasons.

So, if we’re going to proceed with this approach to the binding force of obligation what we need is a sanction that walks softly but also has the proverbial big stick. It must have sufficient heft to serve as a genuine deterrent, but at the same time maintain the capacity to recede into the background, leaving the appropriate, moral, considerations on the table to do the real motivating. If the sanction that I am recommending fits that bill then it will be possible to show how we can be meaningfully bound while still acting for the right reasons.

**3.4 – Sanctions and Exclusionary Reasons**

In place of a transcendental argument for the necessity of the value of rational agency, I start from the position that most of us want to live lives that make sense, and that making sense for creatures like us entails much more than a series of cause and effect explanations. In Section 2.6 I defined an intelligible life as acting in accordance with the principles that together constitute your self-conception. In this section I will expand on that rather basic idea.

We do not simply want it to be the case that for everything we do there is some cause. Rather, we want our lives to make normative sense, which is to say that we want to be able to justify our actions and understand them in the context of all of the prospective reasons that may count for or against those actions. This is not an interest that any of us absolutely has to acknowledge, and it only takes the form of a requirement insofar as we care about living an intelligible life. I think most of us would endorse the principle that we should act on the basis of reasons, but this is not enough to live an intelligible life. We must also act under principles, and I think most of us would endorse that too.

This is a more modest version[[113]](#footnote-113) of the Kantian claim that free agents must formulate maxims for action.[[114]](#footnote-114) An agent is not intelligible to us if he does not deal with like cases in a like fashion. Put another way, you can act for reasons without acting in a consistent or coherent way. In order to forge an identity that persists over time you must commit not only to some particular principles, but to acting under principles in general.[[115]](#footnote-115) This may not be the principle that we care about the most strongly, or spend the most time thinking about, but it is a deep part of our identities in the sense that it takes regulative priority over the other principles that define us.

If you were a blank slate and you wanted to create your identity from scratch you would face two huge challenges. First, you would need to decide what to care about and how to rank your projects and preferences.[[116]](#footnote-116) Second, you would need to find a way to structure your future deliberations so that the principles and projects that define you provide some practical guidance when it comes to determining what is, and what is not, a reason for action.

As I explained in the last chapter,[[117]](#footnote-117) we can adapt Raz’s idea of exclusionary reasons to build an account of how to intelligibly structure our responses to all of the considerations that carry prospective weight in our deliberations. A principle becomes part of our identity when it excludes particular considerations from counting as reasons in certain contexts for us. What gives it force is the cost of erasing that part of one’s self-conception and undermining one’s identity as a whole.

An example might help here. Let’s say that a parent has to buy Christmas presents for her twins. One of them is a model child while the other is usually a little terror. Naturally enough she is more benignly disposed towards the good child, but being a good parent is important to her identity and part of being a good parent is treating your children equally when it comes to buying them presents. This principle rules out certain considerations from counting when it comes to present-buying, one of which being whether or not the parent has a special favourite. Even if she does have a soft spot for one of her twins, come Christmas time that is not a reason for spending extra money on the good one or giving a lump of coal to the naughty one.[[118]](#footnote-118) If she were to ignore that principle and prioritise her favourite then she would undermine her claim to being a good parent and sacrifice that part of her identity.

We incorporate many such principles into the way that we think of ourselves. An integral part of being human is figuring out what relationships and goals we want to invest our time and effort in and then deciding how to fit them all together.[[119]](#footnote-119) This is also true of identities. For most of us, this would, I suspect, be all but impossible if we were not already provided with parts of our identity by our parents and communities.[[120]](#footnote-120) Of course, we have the capacity to reject some parts of our identity and to adopt and fashion new elements as we grow and develop as individuals, but this can only be meaningfully accomplished if it is done in accordance with some existing principles that we have already embraced.[[121]](#footnote-121) Tinkering with one’s identity for the purpose of avoiding a sanction would be contrary to the interest that most of us take in being the sort of person who acts for good reasons and would, therefore, be self-defeating.

The exclusionary principles that we endorse combine to form a hierarchy of commitments, and it is that whole that I mean by a person’s identity here. As beings able to initiate action, we can undermine and damage our identities. And, as in the example of the partial parent above, one particular principle might be compromised. This indirectly affects the whole edifice as it can precipitate a crisis of confidence in one’s ability to resist temptation, in the form of excluded reasons, in other contexts too. Sometimes we may not see how particular actions undermine our self-understanding and in these cases we do not actually experience any penalty for these actions. However, what is important is that the identity that we have constructed is weakened or damaged in a way that we could conceivably come to recognise.

Why else should someone be concerned upon realising, say, that his whole life has been a sham? His ignorance meant that there was no problem at the time and he can take steps to address his shortcomings in the future, but it is very distressing nonetheless to discover that you are not who you thought you are. One reason is that you may have very much liked the self-image that you used to have. Another, deeper, reason is that you are now presented with a question that feels like a problem: “who am I?” For most of us it is unspeakably important to be someone in particular, and that means that we each have to determine ourselves in at least some ways in order to shape an identity. The alternative is something that we are usually not prepared to countenance and that is why the loss of identity provides a weighty sanction that binds us to our commitments.

Consider Rocky. In Rocky III our protagonist is distraught to learn that his well-intentioned trainer Mickey has been secretly arranging it so that Rocky avoids fighting the most dangerous opponents, picking weaker challengers in order to keep Rocky safe[[122]](#footnote-122) and extend his reign as heavyweight champion of the world. Now, Rocky himself has done nothing wrong so it’s not appropriate for him to feel guilty. Neither has he let himself down. He never deliberately or knowingly violated any of his principles so it wouldn’t be right for him to feel ashamed. No matter, Rocky’s a fighter and fighters don’t shirk the tough battles. Learning that he is not the champion and fighter that he thought he was is portrayed as a profoundly painful experience, and the explanation is that his actions do not match his conception of himself.

In this section I have shown how the potency of a concern to maintain one’s identity as an agent living an intelligible life welds exclusionary reasons together into a more or less coherent whole. I turn now to the question with which I began: whether a view that relies on such a cost to generate binding force must reduce morality to self-interest.

**3.5 – Doing the Right Thing for the Right Reasons**

The concern to preserve one’s identity takes a very particular form, which can be described in terms of attention-directedness.[[123]](#footnote-123) Even if we accept that some principles can be empowered to exclude various considerations as legitimate reasons for action, moral deliberation is often a messy process and all manner of possible motivations may flit in and out. It is too much to expect that the existence of a duty precludes contemplation of either inappropriate considerations or impermissible courses of action. What really matters is that the content of an obligation typically dominates the thinking of the bearer of that obligation. Her attention is directed insistently towards the right thing to do.

And is this not exactly what it feels like to be under an obligation? Wouldn’t it be easy to skive off work, or keep the wallet you found, or generally make an exception of yourself if it wasn’t for the little voice that leads you back, time and again, to the straight and narrow? If you are like me then you will be familiar with the power of rationalisation when a bad, but tempting, course of action presents itself. Didn’t you work extra hard yesterday? The owner of the wallet would be lucky even to get just his cards back, and wouldn’t it be fair to get a reward anyway? But if you are like me then you will also know how, for want of a better word, conscience often digs in its heels and we are led again and again to the ‘right’ action and the reasons that support it.

The prospect of losing one’s identity, of being nobody is a frightening one, but it is not fear that usually motivates us to preserve and strengthen our identities. That fear is only the other side of the positive drive to be somebody, and what it primarily does is sharpen our focus on the identities that we want to have. The choice to not be nobody is the very same as the choice to *be* somebody. But that choice can only have meaningful content because of the myriad choices involved in selecting and endorsing the more particular principles that coalesce into a substantial identity. Were I to violate my obligations then I would most likely suffer the consequences, but that is not to say that I will act only out of fear of bearing that cost. Since the question that arises concerns not only what I should do, but also how I will understand myself, the possibility of unspooling my identity leads me to attend to the things that I believe and the features of the kind of person that I want to be.

Practical problems that follow from weakening one’s grip on one’s self-conception also have this effect, although indirectly. Failure to stick to your principles makes it harder to commit yourself again in the future in much the same way as breaking a promise lowers one’s trustworthiness in the eyes of others. Just as it would be difficult to engage with our fellows if we could not ask them to rely on our word, it drastically limits our ability to respond to reasons if we have cause to doubt our ability to commit to principles.[[124]](#footnote-124) In that eventuality it is entirely appropriate to reflect on the kind of person that you have become and how you are doing by those reasons that apply to you.

Sticking with the language of egoism, there is, for almost all of us, a self-interested imperative to care about something. But that is as far as it has to go to get a sanction theory off the ground. We can elect to care about things that are not ourselves and we can choose to sacrifice the more traditional aims, goals, and objectives that feature in narrow accounts of self-interest for the sake of other people and even for abstract ideals.

If the drive to live an intelligible life does not determine a particular form for that life to take, if it is simply one over the course of which a person must be committed to the principle of acting under principles, then there is no cause to think that the relevant reasons that count for and against our possible actions must be self-interested ones. We would be bound by the increasingly concrete and determinate principles that we endorse, and these are selected in the context of our sensitivity to concerns about the value of other people, the badness of suffering, and on, and on. The question then becomes: should we be worried that the foundational self-interested imperative to care about something will fatally undermine our attempts to have a genuinely binding commitment to familiar moral principles? I see no good reason to think that it will.

How much will this help with the other objections connected to self-interest? The trick, I am suggesting, is to ease away from the demand for a categorical morality and rely instead on the fact that most individuals choose to be moral insofar as they endorse as a deep part of their identities principles that enshrine the standing of others. What Cohen does not acknowledge is the difficulty involved in remaking your practical identity. You can decide that you are no longer bound by a law, but a sincere resolution is not the same thing as successfully changing your self-understanding and, thereby, your real commitments. This is what Raskolnikov discovers in *Crime and Punishment*. Our identities are not easily shelved or altered and to act contrary to them in an extreme way is to invite doubt, recrimination, and self-loathing.

This may seem to be a trivial point but, as I have already pointed out, other people figure in the construction and maintenance of our identities. Our self-understanding is, therefore, tied to them in profound and interesting ways. We also typically care about other people and we typically care about what they think of us. What someone else thinks of us can be abstracted a further step to what they should think of us, or what they are entitled to think of us. To understand oneself as a free rider in a moral sense is either to renounce all the nested identities that depend on one’s communal roles, or to make an exception of oneself which is extremely difficult to justify since none of us are exceptional in a relevant way.

We should not expect the binding force of obligation to turn out to be so strong as to make it impossible to do bad things. If that were the case then we would not be free. If it provides a counterweight to our inclinations to do bad things or actions that are contrary to our identities then it is doing its job. And if a theory can account for the phenomenology of bindingness, including the variability of the strength of different obligations, then it will have achieved something useful.

What should a sanction theorist say about the cases when we know that we can do a mildly bad thing and get away with it as far as our own self-image goes? For one thing, it does seem as if we are fooling ourselves. Exceptions weaken our commitments, consider how quickly countless diets and New Year’s resolutions evaporate after the first lapse. For another, consider how rare it is to make an exception like that without a nagging, uncomfortable sense that something is wrong. So long as conscience is pricking and prodding then there is a meaningful sense in which we are bound. That we are not always bound so strongly as to make wrongdoing impossible is not, in itself, much of an objection.

The attention-directedness of the abstract sanction that I am recommending opens up an intriguing possibility for my theory of the binding force of obligation. On my view it might be possible for a person to do her duty cheerfully and incorporate her obligations into her conception of the good life. I will explore this idea next.

**3.6 – Grace**

A key part of the Christian theological tradition is that people are corrupt, sinful creatures who are always inclined to shirk their duties when faced with a choice between right action and satisfying their desires. Emerging from that tradition, Kant was still working with a view of human beings as fundamentally split between a rational self and a sensual self, locked in a permanent struggle for supremacy. The best that could be hoped for was to strengthen and support the thinking self and thereby make it a little easier to keep our desires in check.

Kant (1996, p.164/6:405) contrasted the human condition with holy beings “in whom no hindering impulses would impede the [moral] law of its will and who would thus gladly do everything in conformity with the law”.[[125]](#footnote-125) One onerous demand that this view makes is constant suspicion of our desires and impulses. If duty adopts a pleasant visage then we must beware.[[126]](#footnote-126) We might be inclined to act in what appears to be the correct way because it makes us happy and not because it is the right thing to do. The upshot of this is that we cannot ever completely assimilate our duty into our identities. Obligation must remain a constraint and burden. Anything more becoming than a moral anchor is a siren that hijacks our sensible nature for heteronomous ends.

Of course, not everyone saw things in this way. Friedrich Schiller, for example, sought harmony between the rational and sensible sides of our nature and objected: “[m]ust an unselfish emotion in the noblest of breasts come under suspicion just because impure inclinations often usurp the name of virtue?” (Schiller 2005, p.151). He rejected “the suggestion that in being guided by reason in acting morally, the virtuous person will always undergo some experience of struggle and so will go against their sensuous self, which means that they cannot undertake the action naturally and with ease or joy” (Stern 2012, p.115).

A confident, cheerful inclination to perform morally worthy actions is the state that Schiller describes as grace, and it is not difficult to see why it is a much more attractive prospect than the alternative.[[127]](#footnote-127) It offers the tantalising possibility that we really can become better people, in the sense of integrating our moral responsibilities into a good, enjoyable human life, instead of spending our days wary of our motivations and distrustful of good deeds. This can be achieved by forming our identities around those principles that seem to us best justified. In Section 2.7 I argued that moral approbation should often be directed to a person’s character rather than to her individual actions. When a principle is so firmly embedded in one’s self-conception that to violate it is not only unacceptable, but unthinkable, then it can be a simple thing to express it in action.

As the last section demonstrated, the perspective I am offering is intended to provide a way to reconcile the binding force of obligation with the reasons that we have for moral actions in the first place. In my terminology, to act under principles is to act intelligibly and a failure to stick to those principles with which one identifies by acting on reasons that they exclude makes it more difficult for you to understand yourself as a person living a certain kind of life, and that, I have argued, is a substantial cost to bear.

Identifying with moral principles and incorporating them into the way that you view and value yourself makes grace a possibility. There are two mutually supporting reasons that make this the case. For one thing, we can voluntarily endorse our moral principles because they make sense to us and because we are persuaded by the considerations that count in their favour.[[128]](#footnote-128) Therefore, it is always a possibility that we simply attend to these reasons and so act naturally on that basis. When we have an eye simply on the action in question rather than on what we are supposed to do, Schiller calls this forgetting oneself in one’s actions.

The other reason is that it matters to us that we live an intelligible life, so the sanction of unpicking our self-image takes the form of attention-directedness to the right reasons. The binding force of obligation works by returning our reflective minds over and over again to the right reasons. I have suggested that, for most of us, our interest in living an intelligible life is the deepest interest that we have. If this is right then we need not remain beholden to the assumption that reason and desire are enemies and may instead attempt a synthesis.

This is, of course, not to run before we can walk and claim that grace is easy or inevitable. We might still distinguish between holy wills who never experience contrary inclinations and the rest of us who will often feel the lure of temptation. For Schiller, grace is something that we may experience in rare moments when the whole of our being is pulling in the same direction.[[129]](#footnote-129) The Greeks thought of it as a movable beauty since it “can appear in a subject by chance and disappear in the same way” (Schiller 2005, p.125).[[130]](#footnote-130) We can, perhaps, be a little more optimistic, but it is certainly a state that will not be easy to achieve, and harder still to maintain. It may be best to think of it as a worthy and desirable goal but it is still a significant advantage of my theory that it shows that grace is neither ruled out by definition nor impossible to achieve.

**3.7 – Conclusion**

In this chapter, I have explicated in greater detail the sanction of unspooling one’s practical identity that follows from violating a principle to which one is committed. The main objections to a sanction view revolve around the worry that it reduces moral obligation to self-interest, thereby corrupting our moral reasoning and weakening the ties that are supposed to bind us to moral principles. In response I have argued that even though a sanction is required if we are to speak meaningfully of a requirement, the fear of unpicking one’s identity is a tangible sanction and a robust one.

Not only this, it is no more than the negative form of the deep interest that most of us have in understanding ourselves as the kinds of beings who act on the basis of reasons and, if we want to live intelligible lives, under principles. As such, it works by dragging our attention back, time and again, to the commitments we endorse and the reasons that support them. In this way we can be bound and yet still do the right thing for the right reasons. Finally, I considered our potential for exceeding moral continence and ascending to a state of grace and argued that we can indeed reconcile the binding force of our obligations with our desiring nature, thereby opening up the attractive possibility that we can do our duty cheerfully.

Over the course of the next two chapters I will apply the identity-based sanction theory of obligation that I have developed here and argue that it helps to establish original and fruitful accounts of political legitimacy and individual autonomy. In the first case, it is important to consider how the justification for a state matches up to the identity of its citizens. In the second, binding oneself to a hierarchy of principles is an effective way to respond to reasons and thereby to live an intelligible life.

# 4

# *Legitimacy & Practical Identity*

**4.1 – Introduction**

The dominant view in contemporary political philosophy explicates political legitimacy in terms of moral justification for the wielding of coercive power. It is sometimes thought that moral justification also serves to ground a stronger understanding of legitimacy as the authority of a government to hold, issue, and enforce political obligations. In this chapter I will argue that moral justification should form only one part of our conception of political legitimacy. A legitimate state is a morally justified state where, crucially, that justification matches the practical identity of a citizen. Only then can a state create binding duties for that citizen. The upshot of this is that legitimacy is best understood as a binary relationship that can exist between an individual and her state.

A state’s ability to create binding obligations for its citizens matters for two reasons. First, it is relevant to its moral justification. I have argued that, on my theory of the binding force of obligation, it generally makes a tangible difference to be under a duty. This is primarily because the prospect of undermining one’s identity serves to direct attention back to those considerations that support the existence of the duty. The greater the proportion of citizens who are bound to support and uphold a state and adhere to its directives, the more efficacious that state is likely to be. One of the standard conditions on the justification of a state is that it have the ability to impose its laws on those people who fall within its jurisdiction. Therefore, it is significant if a state possesses the capacity to create new obligations for its citizens.

Second, so long as a state is worth being a member of, it is a good for creatures like ourselves to identify with the structures and institutions of political power because of the contribution that this makes to our ability to live an intelligible life. I will argue that suitably just states can help us to respond collectively to reasons by extending our individual identities around new laws and policies. My conception of legitimacy is sensitive to this dimension of our interaction with political authorities because it explains how a state can play a meaning-giving role, creating and preserving rational social structures that we have reason to support.

I have suggested in previous chapters that a concern for voluntariness is central to the social contract tradition. On my view, it is doubly significant for legitimacy. In addition to another standard condition on the moral justification of coercive political authorities, that space is maintained for individuals to make voluntary choices and that some of those are choices about how their community is run and by whom it is run, I contend that voluntary commitment on the part of citizens to their practical identities can support a voluntary commitment to the state’s ability to tap into the binding force of obligation. When this condition is met, a duty to obey a just state can be autonomous and greatly enhance an agent’s ability to respond to reasons.

A state is legitimate, then, with respect to a particular citizen only if it is justified *and* that justification follows from the citizen’s practical identity. If the two are not compatible then the state may still be fully justified in coercing an individual, but it cannot create *binding* duties for her. In these circumstances, it adds nothing at all to proclaim legitimacy and it takes something away from our

ability to capture the significant relationship that does exist between citizens and their institutions when the citizens’ practical identities are consistent with the justification of their state. This means that if we want to fashion a serviceable normative account of legitimacy we need to incorporate elements of what is often called a ‘descriptive’ approach. Justification, by itself, will not get us what we want.

There are five major advantages to adopting my account of legitimacy. (1) It explains why it matters for a state to be able to impose duties on its citizens. (2) It allows us to make finer-grained judgments about the legitimacy of unpopular political administrations since their low standing, while still problematic, may not affect their ability to impose duties if citizens are sufficiently committed to the system as a whole. (3) Equating political legitimacy with moral justification leads to avoidable conceptual confusion. I argue that my definition contributes to a clearer and more satisfactory overall scheme. (4) It explains the good for creatures like ourselves of identifying with the structures and institutions of a suitably just state in terms of the contribution that this can make to our ability to live an intelligible life. (5) It shows how we can do better than Rawls and Nagel thought possible by realising, rather than merely approximating, the value of voluntariness in the coercive setting of the state.

I will begin by studying Max Weber’s classic statement of descriptive legitimacy and proceed to investigate the distinction between normative and descriptive accounts. Section 4.3 will examine legitimacy in context and explore the relationship between political legitimacy and moral justification. In Section 4.4 I raise the question of the significance of practical identity and explain how political actors can be empowered to extend and determine parts of ordinary citizens’ self-conceptions. I will then turn, in Section 4.5, to consider an important objection raised by Allen Buchanan who doubts that questions of obligation are really relevant to our attempts to understand and justify political power. In response to Buchanan I will argue that it matters a great deal how citizens of a state relate to the principles underpinning the justification of its authority. Section 4.6 will make a further case for tying legitimacy to obligation by outlining the advantages of my view. I will concentrate in Section 4.7 on why it matters that we voluntarily bound to our political obligations. Finally, Section 4.8 will look at Corey Brettschneider’s ideas on the importance of democratic persuasion and consider how my proposals in this chapter might bear on the policy and conduct of states with regards to their citizens.

**4.2 – Normative and Descriptive Legitimacy**

In a celebrated lecture to the Free Students Society in 1919 on ‘Politics as a Vocation’, Weber defined legitimacy as an attitude of support or compliance held by the populace of a state towards their political masters. This support plays a crucial role in Weber’s theory of political organisation because it makes the large-scale use of coercive force possible over an extended period of time: “Like the political organizations that preceded it historically, the state represents a relationship in which people *rule over* other people. This relationship is based on the legitimate use of force (that is to say, force that is perceived to be legitimate). If the state is to survive, those who are ruled over must always *acquiesce* in the authority that is claimed by the rulers of the day” (Weber 2004, p.34, original emphasis). So, legitimacy denotes a particular public sentiment of acquiescence towards the agents and institutions of the state.

Weber described three ‘pure’ sources of authority that could cement the right to rule in the public’s imagination. The first is the authority of the ‘eternal yesterday’ – the psychological grip of the received power structures and existing social mores that constitute a political and cultural tradition. The second he calls ‘charisma’, and by this he intends us to understand the personal reputation of particular leaders and the faith in their ability and character that they engender in a people. Finally, legitimacy can be conferred “by virtue of the belief in the validity of legal *statutes* and practical ‘competence’ based on rational rules” (Weber 2004, p.34, original emphasis). Here it is a belief in the rationality of legal structures that gives rise to a belief in the right to enforce them and the compliance that follows. Weber (2004, p.35) believed that legitimacy was a necessary component of a functional state because continuous administration “requires that human action should be predisposed to obedience toward the rulers who claim to be the agents of legitimate force”. In essence, unless people are prepared to obey, there is no way to manage a modern state in all its complexity. Legitimacy, then, describes a vital feature of the relationship between the public and the institutions and agents of the state.

Here, though, we can see how a descriptive focus on the existence of a particular social phenomenon neglects the more interesting questions that jump out at us. Is it a good idea for some people to submit to other people who claim to have authority over them? Is it not worrying that archaic social traditions could render moral agents politically docile and pliable? What reasons could actually ground a right for some individuals to rule over the rest of us? And we can see how Weber’s use of the concept is naturally superseded by a normative theory in the revisionary Marxism of Jürgen Habermas’s (1984) account of legitimation crises.[[131]](#footnote-131)

Habermas starts by discussing the ideological function of conceptions of legitimacy, whereby citizens are reconciled to the relations of production and political superstructure that dominate a society at a given moment. Legitimation crises arise when the political and cultural apparatus fail to provide a viable ‘life-world’ for individuals, a unified perspective which can impose meaning and order on an otherwise chaotic social world. When economic stresses and political inconsistencies outpace the explanations and justifications that are made available within the life-world, then the shortcomings and contradictions of the economic and political arrangements themselves are uncovered and become the subject of pointed general concern. “The legitimation of orders of authority and basic norms can be understood as a specialization of this ‘meaning-giving’ function” (Habermas 1984, p.118). In this respect, then, a state or a government must be on guard to maintain the positive attitude that Weber described, but this is no longer explained simply in terms of a practical necessity for a political system to function. Instead, a rational justification of power relations is internally required within a viable life-world because the point of the life-world is to provide meaning for individuals. Given the sorts of thinking, reasoning creatures that we are, any claim to the authority to issue and enforce binding duties will eventually have to be backed up by good reasons or it will fail to secure the widespread compliance on which its success depends.

Once we begin to enquire about the role that ‘legitimacy’ plays in our moral and political discourse we can see that a normative understanding is, as Buchanan (2002, p.689) points out, “essential for a descriptive account”.[[132]](#footnote-132) Whether or not enough people are positively disposed to the state and its agents’ claims to wield political power depends on whether there are good reasons to accept the justifications that the state can offer for its possession of that power. A descriptive account of legitimacy describes a quintessentially normative relationship.

Of course, that some citizens accept a justification does not make it a good one. However, this obvious truth sometimes obscures two important points about legitimacy. First, it is also a requirement for legitimacy that a prospective authority can be effective and for that to be the case a sizable proportion of citizens usually have to accept that it has a right to rule.[[133]](#footnote-133) Second, and more pertinently for our purposes here, it is only when the justification of the state is suitably related to the basic values that a citizen endorses that we can meaningfully say that the state can create *binding* duties for her. But I am getting ahead of myself. Having shown that we must aim for a normative account, the next task is to get a little bit clearer about what legitimacy really is.

**4.3 – The Foundations of Legitimacy**

In Section 1.3 I defined a legitimate authority as one that has the right to issue, hold, and enforce obligations. Now that we have an account of obligation we can be more precise. In particular, we can explain how an authority creates binding obligations for its citizens. Talk of legitimacy is intended to capture some important difference between raw power and authority, where submission to the latter is justified and thereby rendered acceptable.[[134]](#footnote-134) Indeed, the history of political theory is the story of the attempt to motivate this distinction. Legitimacy, then, in its most deceptively simple formulation is just this status of being acceptable. But what counts as acceptable, and just who is supposed to be doing the accepting? We can identify three levels, or contexts, where questions of legitimacy arise and which furnish the resources to define it in greater detail. These contexts are legal, political, and moral.

Legal legitimacy can be relatively straightforwardly explicated in terms of the requirement to conform to applicable laws. When your neighbour satisfies the planning requirements for his new extension then we can say that he is entitled to go ahead and build it, his course of action and title to the extension are legitimate. Following due procedure counts as sufficient reason for the rest of us to acknowledge and grant his right to proceed. I will not dwell on this aspect of the concept here.[[135]](#footnote-135) Legal legitimacy is interesting because it is clear that it is not freestanding, it depends on the legitimacy of the system itself.[[136]](#footnote-136)

Whether or not the revolutionary who steals a loaf of bread is right to claim that the law under which she is tried is itself illegitimate is a deeper issue. We can query whether laws are themselves legally (constitutionally) instituted,[[137]](#footnote-137) but we can also sensibly ask the question of whether the system itself is legitimate. At stake here is the acceptability of the established political hierarchy and the system of social institutions and informal structures that come along with it.

In the first place, our revolutionary is likely to appeal to her fellow citizens to cast off their shackles and rise up. As Weber rightly recognised, without widespread compliance and a solid foundation of popular support the modern state is going to find it very difficult to get anything done.[[138]](#footnote-138) From the point of view of efficiency, it matters a great deal whether citizens identify with their state and accept its claim to power. However, it appears that this ‘political’ sense of legitimacy is folded into a deeper, moral picture in a number of ways.

We might draw a comparison between a state that has the support of its citizens, but insufficient justification to rule over them, and one that has both. A natural way to describe the first would be to say that it has ‘political’legitimacy because it can draw on the support of some significant proportion of its people. This is, of course, a key ingredient for any elite that intends to maintain power for very long. If you were a Machiavellian of a particularly realist stripe you might hold that the accumulation of power is the goal of the practice of politics and legitimacy is nothing more than success in maintaining the desired public opinion that contributes to that pursuit.[[139]](#footnote-139) This is mistaken. As we have already seen,[[140]](#footnote-140) a descriptive theory collapses into a normative one, but that is not to say that descriptive elements do not feed into a moral theory of political legitimacy.

In order to be fit for purpose, a state must be able to impose order and make its rules stick, which is why political authorities claim a monopoly over the use of coercive force. If a lack of popular support fatally undermines a leader’s ability to get the job done then she is no longer an authority and should step gracefully aside. While I will discuss stability in detail in Chapter 6, one of my objectives in this chapter is to show how citizens themselves can play a role in securing the efficacy of a just state by incorporating the reasons that support its moral justification into their own self-conceptions, thus binding themselves to do as they are told.

On a sanction view of obligation, the element of restraint is explained by the cost that will follow in the event of a breach. I have proposed a sanction built around the importance for most of us of being the kind of person who responds appropriately to reasons in the course of living an intelligible life.[[141]](#footnote-141) Although we find the prospect of unintelligibility to be unacceptable, the principal effect of this penalty is to steer attention back to the question of who we really want to be and, in the end, to the reasons underlying the duty in the first place. To be committed to a principle in this fashion is not an abstract proposition. If acknowledged, it will shape a person’s deliberations. Even if someone is not aware of how an action will destabilise her identity, it remains significant because she could be brought to see the conflict and would then suffer the sanction.

Unless you are a strong philosophical anarchist you will accept that there could be good reasons to submit to a suitably just state.[[142]](#footnote-142) Indeed, the general consensus is that there are good reasons to support just states. One way of responding to these reasons is to integrate a principle of obedience to a morally justified political authority into your identity. If this is accomplished then you will be tangibly bound to comply with its orders and laws. If large numbers of its citizens are bound to observe its authority in this fashion, then a state will likely be stronger and more effective. The fit between the justification underpinning a state and the identities of its citizens is, therefore, relevant to its moral justification and, ultimately, its legitimacy.

If, on the other hand, a state lacks the capacity to call on the support of enough people then it might lack the capacity to effectively enforce its directives. It may thus fail to be of much use to someone who, in fact, has a corresponding identity. In this case its moral justification for creating duties for her is weakened and may collapse. Further, its moral permission to coerce dissenters depends on its being effective enough to perform the kinds of services on which its general justification is based, so that too can be undermined by an inability to create duties for its citizens if its general effectiveness is sufficiently diminished.

This conclusion can be bolstered by considering another way in which a state’s relationship with its citizens is relevant to its moral justification. Most theories of the state assume that part of what justifies a state’s ability to compel individuals to behave in certain ways is that it guarantees for its citizens large spheres of personal freedom and cedes to them some control over the composition and policy goals of their government so that they can precipitate changes if they are unhappy. A condition of legitimacy is, therefore, that the state actively seeks to treat its citizens with respect, for example by consulting and including them in its decision-making processes.[[143]](#footnote-143)

Morally legitimate states, then, must walk a very fine line between respecting their citizens and standing over them, prepared to force them to comply with directives where necessary, and all the while taking care to cultivate their goodwill. When a high proportion of citizens feel obliged to back the state it will need to rely less heavily on brute force and emergency measures to maintain order and, as a result, will be better placed to treat its citizens with maximal respect.

I will explain in the next section exactly how a political authority can deploy the binding force of obligation. When the justification for a state’s right to issue orders follows from principles through which individuals understand themselves, those individuals empower the state to build upon and thereby extend their practical identities around, say, a new law. Broadly Rawlsian contract theories can, I think, be reassessed along these lines since they start from widely accepted principles and then proceed to derive more substantive content. I shall call the prospective practical identity developed from a contract theory’s first principles a ‘contracting identity’.[[144]](#footnote-144)

A contracting identity is compatible with many richer practical identities, but provides the basic regulative framework within which citizens can then respond to reasons in their own ways. If an authority is to create binding obligations, which it must if it is to be legitimate, then the justification for the authority must dovetail with the practical identity of at least some of its citizens. Those citizens will then be unable to violate their political obligations without undermining their self-understanding.

**4.4 – Extending a Practical Identity**

Say that I am committed to the Rawlsian principle of respect for persons and then grant that he is correct in specifying the principles that broadly follow from that, including the natural duty of justice.[[145]](#footnote-145) What happens to my normative situation when a just state (that exists and applies to me)[[146]](#footnote-146) makes a new law?

The first thing to note is that, so long as the state acts within its remit, it establishes a reason for me to do whatever it specifies in the law just by passing and promulgating it. Indeed, I have what will, in most circumstances, count as conclusive reason to do so. This is because a just state is a Razian authority with regard to me; obeying it helps me to better comply with my reasons overall, especially with what we might call my reasons of justice.[[147]](#footnote-147) The justification for my conformity is derived from my pre-existing reasons, but when a genuine authority gives me an order it changes my normative situation all the same.[[148]](#footnote-148)

I have, for example, good cause to wear a seatbelt when I drive. There are some considerations that count against it: it feels uncomfortable and it restricts my movement. On the balance of reasons, however, these disadvantages are heavily outweighed by the benefits of buckling up. What changes when the government makes it mandatory to wear seatbelts is that it issues an intermediary reason that purports to exclude those first-order considerations that count against wearing my seatbelt.

Of course, Raz thinks that authority works by automatically subsuming and replacing a group of our reasons with a single reason for obedience. This, he thinks, is the essence of obligation. I have argued for something a bit more complicated. If we are to be under an obligation then it must be the case that violation of an exclusionary reason opens us up to the sanction of undermining our self-conception and thereby our intelligibility as a certain kind of human agent.

Now, moral justification can establish liability. It is appropriate for an agent to suffer a sanction if she contravenes a principle and there is a sufficiently strong justification for endorsing and preserving a general practical identity of which that particular principle is a necessary part. For example, people should not break into one another’s houses just to snoop around. If that is right then it would be appropriate for a nosey neighbour who crosses that line to bear a sanction. Justification does not, however, tie a sanction to an individual who does not have a compatible practical identity. The reason for this is that it is ultimately up to the individual to impose a cost on herself in terms of her own self-understanding. If that person’s practical identity is not undermined then she will not have cause to sanction herself in this way and is, therefore, not appropriately bound.[[149]](#footnote-149)

Consider again the state that is not only justified, but justified by my own lights. It will be appropriate for me to suffer a sanction if I flaunt its rules but, crucially, it will also destabilise my self-conception to do so without suitably good, non-excluded reasons. Such a state stands in a special relationship to me. It has the ability to create new rules for me that, other things being equal of course, I must follow on pain of self-censure. This is what I mean by the capacity of a political authority to extend the practical identity of its citizens. Its decisions have a direct effect by specifying new rules that citizens with compatible practical identities must follow if they are not to compromise their identities. My contention here is that we need a theory of political legitimacy that acknowledges the significance of this ability for just states with regard to at least some of the people who are subject to their authority.

There is one very important consequence of this view. The relationships between states and their citizens must be understood on a case-by-case basis. It is entirely possible for a state to be morally justified with respect to some particular individual yet unable to create a binding duty for her. This is nothing to worry about insofar as that state will be entitled to coerce that individual should circumstances require it, but it should be clear that something worth having is missing. I will also argue, in Section 4.7, that a just state plays an important meaning-giving function for a citizen when its justification follows from the principles with which she voluntarily identifies. Putting these ideas together, I submit that legitimacy is best thought of in terms of a binary relationship between a state and those citizens who are bound to it. A state can still be morally justified in coercing dissenters, but it can only ever be legitimate for citizens who possess a compatible practical identity.

We can now see just how the moral justification of a state depends, in part, on its particular relationships with each of its citizens. If it does not have the ability to extend the practical identities of enough people then it may not be able to effectively impose its will. This will undermine its moral justification and, as a corollary, its standing relative to a citizen who might otherwise have been bound to support it. As we noted in the last section, this may further reduce its effectiveness. All of which will also affect its permission to coerce those individuals who do not have compatible practical identities. This is as we should expect. A normatively legitimate state should meet the criteria for descriptive legitimacy as well.

Before I work through the advantages of this view, in Sections 4.6 and 4.7, and consider its implications for political practice, in Section 4.8, I must pause to confront an important objection raised by Buchanan. I will defend the significance of political obligation from his contention that both citizens and states can get along very nicely without any duties of obedience. His mistake, as I see it, is to conceive of political obligations primarily as additional reasons for citizens to obey states that are authoritative for them in any case. When we see that what it means to be under an obligation is to relate to one’s reasons in a special way, underpinned by the sanction of unintelligibility, then it is easier to see why political obligations matter for state legitimacy.

**4.5 – Authority, Authoritativeness, and Legitimacy**

Buchanan (2002) equates political legitimacy with moral justification and claims the Rawlsian mantle in pursuing this line.[[150]](#footnote-150) He (2002, p.689) provides this definition: “an entity has political legitimacy if and only if it is morally justified in wielding political power, where to wield political power is to attempt to exercise a monopoly, within a jurisdiction, in the making, application, and enforcement of laws”. His position is that we can establish the legitimacy of some governing body by providing sufficient reason for it to have the power to coercively impose some suitably constrained system of rules. Making a good enough case for political power is enough to render its existence morally permissible.

The key point for our investigation is that Buchanan thinks that questions about obligation are a sideshow in political philosophy and serve only to detract from the real issues surrounding the normative foundations of states. In a sense, it is a much more sophisticated version of Ladenson’s notion of a justification right[[151]](#footnote-151) since it places most of the emphasis on the question of whether the political actor is right in taking charge: “In other words, political legitimacy is an agent-justification notion, having to do only with the normative sufficiency of the justification for the act of imposing rules, not with whether those upon whom the rules are imposed have obligations to those who impose the rules” (Buchanan 2002, p.695).

Buchanan (2002, pp.268-269) claims that there are four major questions that political philosophers traditionally ask about political power. First, what justifies governments in wielding it? Second, where (if at all) do we get an obligation to obey that government? Third, what makes it the case that the fact that the government issues a rule gives us a reason to comply with that rule? Fourth, under what conditions do we have sufficient reason to comply with such rules? Buchanan skilfully distinguishes these questions, but his main contribution is to cast doubt on the importance of the second question. He argues that if we can justify a government’s claim to hold a monopoly of coercive force and show that we have sufficient reason to not only treat that government as authoritative, but also to comply with its directives in a wide range of cases, then we have everything we need.

Buchanan appeals to the thought that we can have reason to treat experts such as financial advisors and fitness instructors as authorities because they are better placed to help us to comply with our own reasons. When my financial advisor calls me and tells me to put my money in a new investment scheme, I should, presuming he is trustworthy, put my money in the new scheme because he told me to. Similarly, if my fitness instructor tells me to stop struggling with large weights and to do more sets with small ones then that is enough reason to change my workout. However, it is implausible to suggest that we are under a duty to obey either of them. The same might be true of states. The passage of a new bye-law forbidding me to walk on the grass might be reason enough to stick to the path, but perhaps this can be true without it being the case that I am duty-bound to do as the government says. Political obligations are now widely considered to be very difficult to ground, if most of the time we have strong reasons to do what legitimate (justified) governments tell us to do, then perhaps he is right that a preoccupation with obligation is just gumming up the works.

This move rejects the sufficiency thesis described in Chapter 1 in favour of a narrower conception of legitimacy. For Buchanan, a state is legitimate if it is justified in coercively enforcing the law, whether or not those individuals subject to its power are under an obligation to obey it. He is right to think that a state can sometimes permissibly coerce individuals for whom it cannot create binding duties, but Buchanan has a muddled view of obligation and that is why he is so eager to consign the question of political obligation to the philosophical scrapheap.

For a start, he thinks that the attraction of political obligation as a solution to the puzzle of state authority is a chimerical promise to provide a neat answer to questions one and three, and to offer some guidance in identifying the boundaries of question four.[[152]](#footnote-152) This is mistaken. To be under an obligation to an authority is not to have an extra reason to obey that authority. Rather, it is to relate to your reasons for complying with its orders in a special way, where those orders are tied to your self-understanding. Nor is it the case that our being under an obligation to obey a just state is what justifies it in telling us what to do. Rather, its justification provides a necessary condition for us to have an obligation to do as it says. To be sure, the efficacy of a prospective political authority is not independent of the nature of our relationship with it, but this only serves to underline the connection between the moral justification of the state and the origin of our political obligations.

Indeed, it seems that Buchanan has misinterpreted Raz on this point.[[153]](#footnote-153) The difference between an entity’s being authoritative and its having authority is not a further question to the one about how we best comply with our reasons. We have reason to treat experts as genuine authorities for Raz. What makes them authoritative is that some of our pre-existing reasons for listening to them can be encompassed and replaced by an exclusionary reason to simply do as they say. However, the scope of the exclusionary reason to do as my financial advisor says is very small and disappears entirely when we consider some of my other reasons. This is partly because it only applies to financial matters, but it is mostly because when we consider the value of autonomy and the store that we set in making our own decisions it is generally worth retaining the final say on what we do with our money. We might say that while he really is a theoretical authority in financial matters, he is not an all-things-considered practical authority. States claim to be all-things-considered practical authorities. Even though there are very important limits on what states can justifiably tell us to do, they claim the right to make laws that affect all aspects of our lives on the basis of all of our reasons. The distinction between exuding some authoritativeness and being an actual authority in the end turns out to be one of degree.

I have sought to show how it makes a difference to the moral justification of a state if its citizens take its directives to be obligatory. One might object that this misses the point. If I owe a duty to someone in authority then I wrong her by violating it. I do not wrong my advisor when I put all my money in tulips,[[154]](#footnote-154) therefore I cannot be under an obligation to do as she says. Maybe I do not wrong the government when I saunter onto the freshly-cut lawn, even if I have sufficient reason to restrain myself.[[155]](#footnote-155) If so, I am under no obligation, but Buchanan still has everything he needs.[[156]](#footnote-156)

Whether or not somebody is wronged by my action is determined by the reasons that support her claim against me that I refrain from performing that action. It is what supports the judgment that an action wrongs someone that really matters. If there is sufficient reason for me to take the state to be an all-things-considered authority and my self-understanding is such that I will damage it by countermanding one of its orders, then I have an obligation to obey it. If this is right, then what matters is the extent to which a principle of obedience to a just state is justified and what considerations will be excluded for an individual who incorporates it into her identity.

We might, if we wish, look to ways that political obstruction or dissent are thought to wrong others for evidence that it is unacceptable to jeopardise the existence of a just state.[[157]](#footnote-157) Indeed, Buchanan himself thinks that we can owe it our fellow citizens to obey the decisions of a legitimate democratic government.[[158]](#footnote-158) Deciding who, if anyone at all, has been wronged by a violation of the duty to obey a just state seems to me, however, peripheral to the main issue, which remains the significance of a tangible political duty of obedience.

I have argued that we find the prospect of unspooling our identities by reneging on our commitments to be repugnant. I have also explained how this sanction works in times of temptation by directing attention back to our reasons for identifying with a particular principle in the first place. To be under an obligation to comply with the orders of one’s state is, therefore, no mere abstract proposition. It is likely to affect one’s behaviour and this is why it matters for states that they can extend the identities of their citizens as I have suggested and, if necessary, explain to those citizens how its right to rule over them follows from the principles that they themselves hold as part of who they are.[[159]](#footnote-159)

In the next section I will detail some of the advantages of tying our conception of political legitimacy to obligation in this manner. Section 4.7 will then go on to consider the meaning-giving role that a legitimate state can play for its citizens and explain how binding oneself voluntarily to obey its commands can help an individual to respond to reasons and live an intelligible life.

**4.6 – Advantages of a Relational Account Tied to Obligation**

As outlined in Section 4.4, I understand legitimacy as a relational concept. Of course, since we are talking about the political relations in which individuals can stand to one another, legitimacy is already a relational concept. What makes my conception distinct is that it aims to capture the special connection that can exist between an individual and the entity claiming the right to rule over her when she identifies with the justification for its political authority.

To be legitimate, a state must be sufficiently morally justified, as in Buchanan’s account. However, an authority might meet the standard of justification required to permissibly coerce its citizens without being able to get a grip, so to speak, on a particular individual. In this case, justification and legitimacy come apart. The authority may be justified in forcing that individual to conform to its directives, but he or she is not bound to obey its directives in the sense discussed in Chapters 2 and 3. I suggest that authorities are legitimate with regard to individual citizens insofar as they are justified *and* that justification follows from the contracting identities of individual citizens.

Why should this dimension of the interaction between political authorities and the individuals in their care matter if we can justify the possession and use of coercive force? My argument, if successful, helps to accomplish a number of desirable objectives. First, it allows us to explain in much greater detail the way that the descriptive element of legitimacy feeds into a normative theory. As explained in Section 4.3, the ability to extend citizens’ practical identities and thereby harness the binding force of obligation underpins the efficacy of states and is thereby folded into the story of moral justification.

Section 4.5 illustrated how my theory explains the utility of the capacity to create binding duties for authorities. Since the binding force of obligation is constituted by a sanction, it contributes something tangible to the effectiveness of the state. As we saw, Buchanan was misled by the assumption that whether or not a citizen was under a duty to obey his government made a normative difference to his reasons for complying with its orders. It is actually the other way around. Whether or not a citizen is under a duty depends on his reasons for compliance and on how those reasons fit into his practical identity. This matters because the phenomenology of bindingness makes a crucial motivational difference. Even though there is a sanction hovering overhead, it does not create an additional reason for the citizen to do as he is told. Instead, it focuses his attention back on the exclusionary reason for obedience and perhaps even the other elements of his contracting identity that secure and support it.

This is an important advantage for my theory. Political obligations are not motivationally inert conceptual relationships relevant only to the theoretical justification of the state. They are live elements of many citizens’ psychological makeups and contribute substantially, therefore, to the stability and effectiveness of a state.

Second, this approach facilitates finer-grained judgments about the ability of some prospective authority to meet requirements of efficacy and qualify as legitimate. It is clear enough that even a thoroughly worthy political administration is in deep trouble if nobody acknowledges its authority. On my view, it is not just a question of whether public opinion explicitly supports an administration. It is also significant if that government is still in a position to extend the practical identities of its citizens.

This is sometimes the case when an unpopular, though still justified, government holds office in a state with solid institutions and traditions.[[160]](#footnote-160) Although many people may be negatively disposed towards the particular political actors who occupy official roles, they are still bound by their more general allegiance to the system and this permits the office-holders to create new duties – underpinned by the motivational salience of a sanction theory.

Third, it may help to ease some common conceptual confusion if we can give political legitimacy a more specific content than ‘a morally justified authority.’ As our earlier discussion of Weber and Habermas indicated, the term ‘legitimacy’ still carries the legacy of its historical deployment in political theory and these vestigial uses are apt to ensnare us. If philosophers mean to say that some principle or institution stands in need of moral justification, or that being morally justified is a necessary step for the generation of obligations, then that is just what they should say. Dressing justification up as ‘legitimacy’ does little for brevity, and less for clarity. It would be much better to harmonise the descriptive and normative approaches to legitimacy as I have suggested.

Habermas provides a crucial clue to a fourth benefit of my suggestion which I will discuss in the next section where I look again at Nagel’s (1991, p.36) questionable assertion that “the search for legitimacy can be thought of as an attempt to realize some of the values of voluntary participation, in a system of institutions that is unavoidably compulsory”. My claim will be that my theory of legitimacy is able to account for the value of identifying with the principles behind the political institutions that dominate our social world. I will go on to argue that citizens can be voluntarily committed to support their state in an important way, and that when this is the case it can assist them in living intelligible lives by helping them respond collectively to reasons. Defining legitimacy with reference to a state’s ability to create binding obligations for specific citizens acknowledges the significance of this relationship.

**4.7 – Legitimacy and Voluntariness**

Our membership of a political community that is characterised by a state monopoly on coercive force means that the possibility of living an intelligible life is not independent of the task of justifying political authority. Habermas sees that states play a meaning-giving role because the entire life-world, the entire social order, is organised around the state’s claim to hold, issue, and enforce duties. This must be defended, and defended well. A person who is subject to this kind of practical authority is likely to find it troubling and distressing if she thinks that her political institutions do not have a rational basis. Her life is likely to go better if her practical identity is in harmony with the justification of her political institutions. Identifying to some degree with one’s state contributes to a satisfying life and is, other things being equal, an important good if it can be achieved. My conception of legitimacy is sensitive to the value for individuals of states that fit with their practical identities.

This is not just a question of reconciling an individual to the institutions and structures of power that comprise her state. In the next chapter I will argue that identifying with, and thus binding oneself to, principles is an important expression of autonomy and is partly constitutive of living an intelligible life. As such, it is critical that our states help us to respond to reasons well when we grant them the power to extend our identities. For this reason, it matters that a principle of obedience to just states can be endorsed voluntarily.

It also matters because states and communities work hard to inculcate political values and commitments in citizens from a young age. Sometimes we identify with principles not because we are convinced of their merits but because they have been forced on us in one way or another. This is not always a bad thing. Indeed, part of growing up is picking and choosing between the identities that one has been brought up with and finding ways to fit them together into a coherent and sensible whole.[[161]](#footnote-161) However, an identity that one is committed to because the alternative is unacceptable is not an identity to which one is dedicated as a means of responding to reasons.[[162]](#footnote-162) An identity that is not embraced because of the value that one perceives in it is not one that can provide the right kind of framework for an intelligible life.

Imagine the denizens of a totalitarian regime. Omnipresent propaganda, surveillance, and censorship will have an insidious effect. If successful it will induce its targets to internalise the oppressive rules of the society so that they will censure themselves. It will establish the figurative policeman inside their heads. The absence of personal freedoms and basic political liberties, and their subsequent subjection to the will of the rulers, would be deeply problematic from the point of view of individuals with the capacity to be autonomous agents. However, there is another problem. Individuals will feel bound to obey the commands of the rulers, but relating to the rulers in this way will be just one more element of their lives that will fail to stand up to rational scrutiny should they come to reflect upon it.

I have emphasised here the importance of responding to reasons and living an intelligible life. A possible objection at this point is that it is a perfectly rational and intelligible response to perform an action because it is unacceptable not to do so, especially on a sanction theory since it explains that unacceptability in terms of a tangible cost. It can, therefore, be rational to obey the totalitarian state if an act of disobedience would undermine one’s self-understanding. This is quite correct but it is only an objection on an unhelpfully narrow view of what it means to respond to reasons. Not only do we have an interest in responding to reasons from instant to instant, it also matters to us that the principles in which our identities are bound up are themselves *responses* to reasons and this is not the case if they are forced upon us.

Rawls and Nagel assume that the best we can do is to show how the justification for a state can be held in reflective equilibrium, when our considered judgments and our reasoning from first principles together vindicate a basic structure that unavoidably stands in a coercive relation to those people who live within it. Since this justification would speak to us if only we could think about it in the right way, they conclude that subjection to political authority can therefore be autonomous. This, along with provision of the political liberties, is what it means for them to realise some of the values of voluntary agency in a necessarily coercive setting, rather than the ideal of voluntariness itself.

In contrast, I have contended that our foundational commitments can themselves be substantially voluntary if they are endorsed because of the reasons that count in their favour. If this is right, then a legitimate state, as I define it, can be one in which a citizen voluntarily polices her own political obligations, even as it is up to her government to set some of their content. A state that is legitimate for me is ideally one to which I am voluntarily committed. This is, I think, a more satisfying conceptual scheme than one built only on moral justification because it explains the intuition that voluntary participation is central to the legitimacy of any political authority.

In this section I have argued that my account of legitimacy is sensitive to the meaning-giving function of states that Habermas identified. It is an important good for people to feel that their political authorities are not alien or oppressive. Beyond this, I have also argued that a state is legitimate for an individual when she is voluntarily committed to the principles that buttress the justification of its right to issue orders because of the reasons that count in favour of its having authority. It is only when this is the case that a state can satisfactorily help an individual to respond to her reasons.

What does this mean, if anything at all, for public policy? In the next section I will consider and adapt some of Brettschneider’s (2012) interesting ideas about democratic persuasion to paint, in very broad strokes, a picture of how states might be guided by the conclusions reached in this chapter.

**4.8 – Democratic Persuasion**

Starting from the premise that persons should be respected as free and equal, Brettschneider shares my view that the state can be justified. Since states, ideally, exist to guarantee conditions under which individuals can develop and exercise what Rawls calls their two moral powers,[[163]](#footnote-163) they have a responsibility to promote the values that they are designed to express. States are not, therefore, value neutral in the sense that they are have no values of their own and merely serve to impartially defend a maximal set of liberties for all of their citizens. For Brettschneider, this means that we should advocate what he (2012, p.4) calls ‘value democracy,’ the view that the state “should engage in democratic persuasion, actively defending the democratic values of freedom and equality for all citizens”.

His main aim is to find a way to stand firm on the liberal principle of free speech, as he (2012, p.37) believes that this “right gives citizens an entitlement to say and believe whatever they wish”, without conceding that a liberal state is impotent to resist the spread of hateful doctrines that deny the freedom and equality of all citizens.[[164]](#footnote-164) The solution is to distinguish between the coercive and expressive roles of the state. Brettschneider argues that although the state must permit the expression of discriminatory and objectionable beliefs, it has a responsibility to articulate its foundational principles, rebut hateful viewpoints, and, ultimately, to persuade citizens to adopt its values as their own. This seems right, although I think that we can give democratic persuasion a firmer foundation now that we have a theory of legitimacy that incorporates my theory of the binding force of obligation. Further, we can be more precise about how it must proceed and what its limitations will be.

Democratic persuasion can serve legitimacy in two ways. It can promote stability and efficacy if it leads more citizens to experience the binding force of obligation in relation to the directives of a just state. It can also establish obedience to the state as an autonomous response to reasons for individual citizens and secure for them its meaning-giving role in creating and maintaining rational institutions and norms.

The way to proceed is for the state to find ways to make citizens’ own commitments transparent to them. This is easier said than done. Brettschneider (2012, pp.81-82) cites familiar examples of reasoning intended for public consumption such as the publication of judges’ decisions and more novel possibilities such as dedicating public holidays to men and women whose principles and behaviour exemplify the values of the state.[[165]](#footnote-165) Although these suggestions are helpful and can doubtless form part of the solution, the state should actively seek ways to engage with its citizens as individuals. If citizens can be assisted in drawing the connections between the principles with which they personally identify and the justification of the state then they will quite naturally feel bound to adhere to its rules when it acts justly.

One key way that the state can engage in democratic persuasion is through education, particularly of children and young people. Here the state will have to start by encouraging children to form identities that include a regulative commitment to the principles underpinning the justification of the state. Given the nature and development of children this will necessarily involve a degree of compulsion. However, the goal should not be to produce obedient but unquestioning citizens. Rather, it should be to foster the development of a critical spirit so that individuals can come to voluntarily endorse the values of the state. Only then can the authority of the state form part of their attempts to live autonomous lives as an acceptable response to reasons.

As a child’s education progresses the curriculum should change too and teachers should strive to engage their pupils as rational agents, as indeed many of them already do. There is also no good reason why civic education should suddenly stop at a particular age. If we are to take seriously the liberal exhortation that the justification of the state should be available to all then we must be prepared to invest heavily in facilitating access to education for all citizens.[[166]](#footnote-166)

In the end, it is for individuals themselves to identify with those principles that seem acceptable to them and to police their own obligations. There are limits, therefore, to what democratic persuasion can achieve and how it may be pursued.[[167]](#footnote-167) States also have a moral responsibility to treat their citizens with respect and so must resist any temptation to engage in overbearing or oppressive behaviour for the sake of persuasion.[[168]](#footnote-168) I will once again emphasise that having a contracting identity does not preclude an individual from having and pursuing a unique conception of the good. It does not determine all of her values or specify what she must do or refrain from doing in all circumstances. Its main role is to regulate her other commitments so that she does not allow, say, considerations from her religious beliefs to inappropriately affect her actions in the public sphere. To have a contracting identity, in my account, is to be bound to treat others as free and equal and to support and uphold just states that secure and protect the conditions under which individuals can cultivate and exercise their two moral powers.

**4.9 – Conclusion**

The major claim in this chapter has been that we should understand legitimacy in terms of a binary relationship that can exist between citizens and their states when a state is morally justified and that justification follows from the principles that make up the self-conception of a citizen. I have argued that a legitimate state must be morally justified but that moral justification is not sufficient to secure the efficacy that is itself required as a condition of justification. Morally justified states can impose duties on at least some of their citizens, and thus call upon the binding force of obligation, when the justification of the state follows from the principles that those citizens hold as part of their practical identities. Normative ascriptions of legitimacy, therefore, should include a traditionally descriptive element: how citizens relate to their state.

I responded to Buchanan’s objection that obligation is largely irrelevant to the operation of political actors and the deliberation of citizens by showing how he held a mistaken view of obligation and failed to grasp why it matters that individuals relate to reasons in this special way. I went on to argue that my view facilitates greater precision when it comes to determining the legitimacy of unpopular states and that it can avoid conceptual confusion by bridging the gap between normative and descriptive approaches.

Following Habermas, I acknowledged the meaning-giving role that states play for their citizens and argued that my account of legitimacy is not only sensitive to this feature of legitimacy, but in fact captures it. Identifying with the principles underpinning the authority of a just state can be an effective way of autonomously responding to reasons.

I then considered how my proposals might make a difference to how states conduct themselves with regard to their citizens by considering Brettschneider’s theories of value democracy and democratic persuasion. Legitimate states will work to find ways to address their citizens as individuals and to make the commitments that follow from their own identities transparent.

I have placed a great deal of emphasis on the importance of living an intelligible life and responding to reasons. I have explained these ideas through the capacity to build and maintain a hierarchy of principles that together constitute a practical identity, a self-conception that matters a great deal to most of us. In the next chapter I will rely on these ideas to explicate a substantive theory of autonomy.

# 5

# *Autonomy & Living an Intelligible Life*

**5.1 – Introduction**

You might be tempted to think that political philosophy needs another theory of autonomy like it needs another colourful metaphor: not at all, but what harm could it do? However, I think that my discussion up to this point provides the building blocks for an account of how we go about living autonomous lives and why we value autonomy under a range of descriptions. This is helpful because it allows me to say something more substantial about the interest I have posited in living an intelligible life and it serves to draw out the significance of realising the ideal of voluntariness.

In this chapter I will argue that to live an autonomous life is to respond to reasons by voluntarily binding oneself to a hierarchy of exclusionary principles. To be autonomous in this sense, one must have not only the capacity to commit to principles, but also actually value the possession of an intelligible self-conception. This interest is what underpins the existence of a self-imposed sanction and, thereby, the phenomenon of bindingness. It is only by binding ourselves in this way that our responses to reasons combine to create unique, concrete individual identities. I will, therefore, expound a substantive theory of autonomy that goes significantly beyond conceptions that rely only on the procedural conditions required for a sentient being to be sensitive to normative considerations.

My view relies on the value for creatures like ourselves of living an intelligible life. I will use this idea to explain the underlying unity of three common conceptions of autonomy that are relevant to our moral and political deliberations. The first kind of autonomy concerns the common-sense notion of being able to do what we want. While there are, of course, limits, we have an interest in being in control of our own lives. The second kind comprises a subjective sense of identity with the various social rules and coercive institutions to which we are subject. Ideally, the claims that our political authorities make on us should be justified in such a way that we can come to see the political environment we inhabit as a rational one in which we can feel at home. Finally, there is an even more abstract sense of autonomy that corresponds to an ideal of rationality, under which we strive to respond appropriately to reasons rather than simply satisfying the desires that we happen to have or pursuing the interests that we elect to acknowledge. But in what respect are these ideas the same such that we can bring them all under a single concept, and, indeed, how can we explain their distinctive and apparently independent values if we are successful? I will argue that the significance of reason-responsiveness is the key to appreciating their underlying unity and that they are three dimensions of striving to live an intelligible life. Further, understanding them in this way provides some guidance when it comes to balancing their sometimes competing claims in a political context.

In fact, coercive political authorities help us to live an autonomous life because they facilitate responses to reasons that would not be possible for individuals acting alone. Indeed, I argued in Chapter 4 that part of the point of states is to promote and preserve a more comprehensible and rewarding social environment. Understanding oneself through one’s political responsibilities can

thus be a response to reasons and so political obligation can be an extension of, rather than a constraint upon, autonomy. It matters that these commitments are voluntary ones primarily because of how values figure in our responses to reasons. My contention will be that valuing is a typically voluntary relation towards something and it is values that we seek to instantiate in those principles with which we identify, and according to which we order our responses to reasons.

Procedural accounts of autonomy place great emphasis on the integrity of an agent’s commitments. For example, they often rely on some notion of reflective endorsement to ensure that they are really her own. I will show that reflective endorsement of a principle does not entail commitment for the reason that it does not, without an attendant sanction, generate any binding force. This is why I think that purely procedural accounts of autonomy are missing something important. They lack the crucial element required to secure a method of determining oneself that goes beyond isolated, fragmentary desires, choices, and actions. They lack an account of the binding force required to form an autonomous identity and thereby live an intelligible life.

I will begin by outlining the three conceptions of autonomy I have identified in more detail, before turning in Section 5.3 to address the question of why we care about autonomy under these descriptions and consider the worry that they are distinct values that have been unfortunately (and unhelpfully) lumped together. I will then examine each conception in more detail to show how it connects up with the other conceptions and why it constitutes an integral part of an intelligible life. Section 5.7 will return to the ideal of voluntariness that was first raised in Chapter 1 and explain the significance of realising the ideal itself, rather than some of the values of voluntary participation, in the context of an intelligible life. Finally, in Section 5.8 I will make the case that procedural accounts of autonomy are deficient because they lack the focus on bindingness that I have been pushing in this dissertation.

**5.2 – Three Kinds of Autonomy**

There is, of course, no shortage of definitions of autonomy available in the literature.[[169]](#footnote-169) My objective in this chapter is not to defend my account as a superior conceptual analysis of ‘autonomy’. Rather, I want to flesh out what it means to live what I have described as an ‘intelligible life’. In the process, I will address Flint Schier’s (1993) worry that accounts of autonomy mistakenly tend to wrap up a number of different values that we loosely associate with freedom and agency under a single concept. My conception will stand or fall as a theory of how these values fit together and why they matter to us. I believe that we can divide them into three central themes and these will form the basis of my discussion.[[170]](#footnote-170) I will briefly sketch each of these themes in this section before I explain Schier’s worry in Section 5.3. My strategy then will be to explain how they do, in fact, fit together as constitutive parts of a life lived in response to reasons. When that foundation is in place I will be able to explain the significance of voluntariness.

The most familiar theme revolves around the idea of personal autonomy, which obtains when an individual is in control of her own life. An agent is autonomous, in this sense, when she is free to act as she chooses.[[171]](#footnote-171) For this to be the case two initial conditions must hold. The first is that other people do not interfere with her actions. Interference, whether it comes in the form of straightforward physical coercion or more or less subtle attempts to manipulate the relationship in which an agent stands to her options, hijacks that control and subjects the agent to another person’s will. The second has to do with how an agent goes about formulating her goals[[172]](#footnote-172) and values.[[173]](#footnote-173) In order for any agent to be in control of herself, she must also be substantially free of coercion and manipulation in terms of the mental processes that underpin the selection of aims and objectives and the background evaluations that give them meaning.

We have an interest in deciding for ourselves what we will think and do and the value of personal autonomy provides the basis for claims we make against each other and against institutions of the state for important freedoms such as bodily integrity and freedom of conscience.[[174]](#footnote-174) When an individual’s personal autonomy is sufficiently respected in laws and practices then there is a clear sense in which we can say that she governs her own life because she is accorded a sphere of agency in which she is not subject to the will of others, either in what she does or in how she comes to decide what to do.[[175]](#footnote-175)

Of course, this is subject to further qualifications. It is not difficult to identify examples of psychological afflictions that cannot be explained as a direct case of external interference nor as malevolent manipulation behind the scenes. Think of extreme phobias. These, and other similar conditions, undermine personal autonomy in so far as it is implausible to hold that an agent is in control of her actions.[[176]](#footnote-176) We should also note the familiar, but nonetheless profoundly important point that there are limits to the actions that we are permitted to perform, imposed for the sake of preserving the personal autonomy of others,[[177]](#footnote-177) and to promote other goods such as welfare, excellence, and justice.

The value of personal autonomy seems so obvious that it is easy to take it for granted, but it is worth pausing for a moment to consider why we hold it to be so important. Two possibilities immediately present themselves. First, personal autonomy is valuable because it is partly constitutive of a good life for a human being to make choices and thereby become “efficacious in the world,” as Thomas Hurka (2011, p.149) suggests. Or, second, personal autonomy is valuable because without it we could not pursue our particular conceptions of the good.[[178]](#footnote-178)

Isaiah Berlin (1969) famously defended ‘negative liberty’ as both an intrinsic and an instrumental good. There will be much more to say about this in Section 5.4, where I will defend a version of the first claim. However, it is worth noting that on either line the value of particular instances of personal autonomy must be understood in a broad context, i.e. a human life. John Christman (1991a, pp.2-4) prefers to concentrate on personal autonomy as a property of individual actions. I will reject this approach here because one of the things that I am trying to establish is why personal autonomy matters to us, and it will emerge that it matters to most of us in the context of an interest we take in living an intelligible life.

I would now like to turn to a less intuitive, though extremely influential sense of autonomy that I broached in the last chapter on legitimacy.[[179]](#footnote-179) The limits that we collectively impose on our range of possible actions and the demands we make of each other take on a greater significance when they fall under the auspices of the state and come with the threat of coercive force. It is one of the principal aims of political philosophy to reconcile the existence of state power with our status as agents with an interest in personal autonomy.

From the perspective of the individual, it matters a great deal whether the structures of the social world make sense and whether there are good reasons for submitting to the authority of the particular people who claim the right to wield power over us. An important form of autonomy is endorsing and identifying with the normative architecture of our communities and states. We can explain this sense of autonomy with the aid of Hegel’s notion of subjective freedom.[[180]](#footnote-180) Dudley Knowles (2002, pp.74-75) describes it as comprehension “of the existent social world as rational,” where the “individual recognizes the validity of substantial ethical norms, identifying with those institutions and their constitutive ethical principles”. In more concrete terms, it does not feel like such a burden to pay your taxes if you think that the government is entitled to collect them and will spend them wisely on worthy causes. Leaving aside questions of justice, it should also be apparent that a subjective sense of identification with the organs of state power is likely to make one’s life go better insofar as an agent does not think himself constrained and oppressed by alien, and perhaps hostile, forces.

Further, binding oneself to a worthy authority is one way of helping to make the world a better place. There are some problems that require collective solutions. I can, for example, try to combat climate change on my own but, needless to say, I will not be successful. Ideally, this is where states come in by coordinating our activities and deploying coercive force where appropriate. A justified state can assist its citizens in responding to reasons but it needs their support. When citizens identify with the authority invested in public institutions, then everyone can reap rewards in terms of stability and efficiency.[[181]](#footnote-181)

However, there is much more to say about the connection between autonomy and reason, and this pushes us on towards our third theme. We might hold that an important element of autonomy is the proper functioning of our rational faculties such that we can respond appropriately to the reasons that apply to us.[[182]](#footnote-182) This is importantly different both from the freedom from extreme aversions and mental illnesses that can undermine personal autonomy and the Hegelian conception of subjective freedom as adherence to principles that you take to be sufficiently justified, in that it corresponds to the value that we place on the ability to reason well and to get things right when we do.

The idea is that a very particular type of freedom consists in liberation from our limitations as embodied beings, determined by our natural drives and desires. Autonomy in this sense is contrasted with heteronomy and value is located in the exercise of our rational faculties. Kant famously grounded all value in the capacity for rational agency. We do not need to take that step in order to think that personal autonomy “is especially valuable if it follows deliberation, and chooses among options that are fully understood” (Hurka 2011, p.149).[[183]](#footnote-183) Nor do we need to take that step to hold that there is something specially worthwhile for reasoning beings like us in justifying to ourselves our social and political structures, and working to perfect them so that we may feel at home within them.[[184]](#footnote-184) Sections 5.4, 5.5, and 5.6 will spend more time getting to grips with the idea of responding to reasons and explain in more detail how we go about creating and conducting an intelligible life. Before that, we must confront the possibility that we are on a fool’s errand. Perhaps it is a mistake to attempt to unify these values in a single account of autonomy.

**5.3 – The Value of Autonomy**

Schier attacked the cosy consensus that has formed around the value of autonomy along two interesting lines.[[185]](#footnote-185) First, focusing on the first and third senses of autonomy outlined in the previous section, he questioned whether they were the same kind of good for human beings. If not, then why should we think that they fall under the same concept? Second, he denied the supposed pre-eminence of responsiveness to reason and the expression of a uniquely rational nature, assigning a primarily instrumental value to our critical faculties. Following Hume, he suggested that the ability to reason well is desirable as a means to achieving ends that are not themselves set by the unconditional value of reason. Reasoning well is a particular kind of excellence, but aside from the perfectionist thought that we do well to cultivate excellence in general, reason matters to us most because it enables us to live smarter, more successful lives. Responding to these challenges will help to clarify the value for almost all human beings in living an intelligible life and demonstrate how the three senses of autonomy with which I began are three pieces that fit neatly together in a particularly desirable form of human life.

Schier starts by identifying two major kinds of unfreedom. Following Aristotle, he distinguishes between unfreedom that can be attributed to the interference of others and the “unfreedom or slavery that consists in some state of the slave’s soul” (Schier 1993, p.1). He goes on to suggest that “this distinction has important consequences for assessing the credibility of any form of liberal political theory that bases itself on the notion of autonomy”. His contention is, of course, that these types of unfreedom are so different that we cannot explain them adequately using a single concept, and, therefore, that any system of thought that takes such a concept as either basic or integral is erected on shaky foundations. Schier’s target is the Kantian liberal tradition founded on the idea that human beings are, to borrow Rawls’s phrase, free and equal rational beings. Autonomy, for Schier, consists solely in the individual ability to reason well, and this is just not the same thing as being able to act as one wishes.

This sundering is complete because he doesn’t just think that negative freedom and autonomy are different, but interdependent, stages of the same value. Instead he takes the significance of each to be independent of the other. So, it is not the case that imprisonment and maltreatment are bad because they undermine autonomy, although that may provide another reason why they are bad. In the first place they are bad because they are contrary to our interest in maintaining an inviolable sphere of personal independence. Were we to locate the significance of negative liberty in the necessary conditions for the existence or emergence of autonomy, Schier thinks that we would we both wrong and foolish: wrong since exceptional individuals may have unconstrained wills in spite of oppression and incarceration;[[186]](#footnote-186) and foolish because it might justify all sorts of restrictions and violations of negative freedom. “So not only might autonomy flourish under the most illiberal regime – the Kantian gulag – but the most illiberal measures could be justified by claiming that they are needed to promote autonomy” (Schier 1993, p.4).

Schier was arguing against the idea that negative freedom is either a form of autonomy, or is primarily justified instrumentally on its behalf. As such, he (1993, p.10) contends that “negative freedom is an irreducible good”, something that we value “as an end in itself, an indispensable component of a worthwhile human existence” (1993, p.14). However, he (1993, p.11) expands on Berlin’s formulation by pointing out that we “do not value freedom from the interference of others only in those spheres where we propose to act ourselves”. We have an interest in ring-fencing such things as control over our sensory access to the world and bodily integrity and this is most plausibly described simply as freedom from interference, as opposed to freedom from interference for the sake of doing something.[[187]](#footnote-187)

We are, according to Schier (1993, p.11), social and territorial animals, and so we take a “natural interest in the protection of certain domains of our life from the meddling of others”.[[188]](#footnote-188) The upshot of all this is that, for Schier (1993, p.11), the value of negative freedom for human beings is ultimately the same as it is for “cats, dogs and other animals”. I will argue against this analogy in the next section.

So what about the unfreedom that consists in a defective or an enfeebled soul? Why is the ability to reason well valuable and how does this value differ from that of negative liberty on Schier’s account? He (1993, p.10) writes that rational autonomy is “something we value because we value excellent judgment and action”. Its intrinsic value is to be understood in perfectionist terms, but reasoning well is a virtue that we hold especially close to our hearts because it enables us to pursue our projects more effectively. Insofar as our capacity to respond to reasons is valuable independently of this instrumental role, Schier maintains that it does not provide a direct reason to care about personal autonomy.

I think Schier is mistaken here. There is a deep connection between expressing our nature as reasoning beings and having some measure of control over our own lives. My aim over the course of the next three sections is to reveal that connection by considering why it matters to us to have personal autonomy, to identify with our states, and to master those drives and impulses that would render us heteronomous.

**5.4 – Personal Autonomy**

Here is the basic idea. Living an intelligible life requires some space within which to make one’s own choices and to shape oneself and the world according to one’s will, in line with the reasons that one finds persuasive.[[189]](#footnote-189) This is better achieved within a social world that is itself comprehensibly ordered and generally predictable.[[190]](#footnote-190) And all of this is made both possible and desirable by the interest we take in responding to reasons. The quality of such a life is progressively improved by cultivating our rational faculties so that we not only reason, but reason well. The three senses of autonomy combine to form a much more complete and attractive picture of a good human life. But what holds them together is the interest in living an intelligible life that I first raised in Chapter 2. I argued there that nearly all of us care very deeply about living a unified life, ordered according to those principles that seem to us best justified. These principles play an exclusionary role, backed up by the cost that follows from a breakdown in our ability to understand ourselves in these terms. Although it is profoundly unsettling to unpick one’s self-image, the good news is that we can choose our particular identities. We can voluntarily commit to the values and principles that constitute our self-conceptions because of the considerations that commend them to us. The presence of a sanction binds us in a tangible way, but we develop our practical identities in the first place because of the reasons that count in their favour. In this section and the two sections that follow I will consider each of the three kinds of autonomy as an element of an intelligible life and show in more detail how they link together.

So why do I aspire to have control over a more or less extensive sphere of agency? Well, there are actually two distinct sets of reasons for why this should matter. One set has to do with the first and second order reasons on the basis of which we ordinarily deliberate and make decisions. Why should it be up to me which drink I order at the bar? Well, that way I’m more likely to get what I want. Why should it be up to me to decide what I wear? Well, I want to make sure that I wear something that I find comfortable. Or perhaps I want to ensure that I am presenting myself to others as an aficionado of the blackest, most brutal death metal out there and I can’t do that without just the right official band t-shirt. In this regard, Schier is entitled to draw parallels between human beings and animals since it is the value of being able to pursue and secure those things that appear desirable to us that grounds a similarly territorial interest in carving out and dominating some personal fiefdom.[[191]](#footnote-191)

There is, however, another set of reasons in play here. It also matters to me that I engage in the world in a practical way: that I *do* some things, and that the things that I do can be justified. Agents not only act, they act for reasons. In this sense my actions acquire a reflexive dimension. They are partly about me.[[192]](#footnote-192)

Reason-responsiveness is key here. You must have some personal autonomy in order to respond appropriately to certain kinds of reasons, which I shall (unimaginatively) call practical reasons. That a cold beer would taste particularly nice on this warm summer’s day is not only a reason to form a belief about what would be pleasant – it is a reason to act. It is a reason to drink a cold beer on a sunny day. Living an intelligible life is not a purely intellectual exercise. Some reasons are practical reasons, which is to say that they are reasons for or against a particular course of action.

I am not just referring here to what Raz (1986, p.145) calls ‘action reasons’. Action reasons are reasons not just for bringing something about, but for performing the relevant action yourself. That some music would liven up this party is not directly a reason for me to whip out a guitar. I could just turn on the stereo. If I had been hired as the evening’s entertainment then I might have an action reason to play the music myself. Raz’s action reasons, then, are just one species of practical reasons.

I shall understand any consideration that counts for or against an action in some situation as a practical reason.[[193]](#footnote-193) Practical reasons mean that one cannot live an intelligible life as a mere spectator. If you do not interact with practical reasons as you encounter them then there will be one very important part of the world that will make no sense to you at all and that, of course, will be yourself. It would be like watching your own life as if it were a film.

Often an individual will feel that all things considered he should not act, but this is a determination made on the basis of the balance of the practical reasons as it appears to him. Sometimes he will wish to act but refrain because he expects to fail and make the situation worse, or because he thinks that someone else will do a better job. Notice that these are also cases where a determination is made on the basis of the practical reasons that an agent recognises. Inaction can be an appropriate response to one’s practical reasons, but inaction requires a measure of personal autonomy too. When chained and masked, Hannibal Lecter is not responding to reason by not eating his guards’ livers with fava beans and a nice Chianti.

This, I think, is why the Ludovico Technique in *A Clockwork Orange* is so disturbing. After the aversion therapy, Alex associates his favourite piece of music, Beethoven’s Ninth Symphony, with the horror he has experienced and finds it intolerable. He has been stripped of the capacity to respond to a beautiful piece of music as its merits dictate. Even though he remembers his fondness for Beethoven, and presumably still considers the Ninth Symphony to be a masterpiece, he is reduced to a whimpering wreck when it is played and is driven to jump through a window rather than listen to it for any length of time.

Without a measure of personal autonomy it would also be very difficult, and perhaps even impossible, to constitute our conceptions of ourselves by discovering what things we like, what things we dislike, and how best to respond to them. We require a degree of personal space to explore the world and to react appropriately to what we find in it. This also explains why personal autonomy cannot be reduced to negative liberty. All kinds of education, training, and resources are required to act on many of the reasons that we encounter as we experience everything our natural and social environments have to offer.

Unlike Schier, I think that some measure of personal autonomy really is a precondition for expressing one’s rational nature since it must be present in order to appropriately respond to those reasons that weigh for or against particular courses of action. This does not mean that personal autonomy has only derivative value. It is a constitutive element in an intelligible life and therefore shares in the value of that uniquely human manner of existence. We develop much of our selves in our actions, some of which we perform deliberately and some of which we retrospectively endorse on reflection. As practical beings it would be a mistake to attempt to understand ourselves apart from our active engagement with the world. One of the most important ways that we can respond to reasons is by conceiving of our own selves in particular ways and thereby limiting our personal autonomy, and this leads me to consider the value of autonomy as identification with the rules of our political communities.

**5.5 – Autonomy as Identification with the State**

States feature in our attempts to live intelligible lives in two ways. First, they play a negative role insofar as they must contribute to a comprehensible and predictable social order if they are not to frustrate and alienate us as we exercise our personal autonomy. On this level, identification with the state means that one does not feel oppressed or dominated, or even just confounded. Second, states positively contribute to our efforts to respond to reasons when they successfully direct collective action and resources to solve large or otherwise intractable problems. Identifying with the justification of the state and its ability to extend your practical identity, therefore, serves to enhance your ability to live an intelligible life. I will deal with these ideas in turn.

Recall what we said about the meaning-giving role of states in the last chapter. Sooner or later citizens demand good reasons to support their coercive political institutions. And of course they are right to do so since one of the functions that governments perform is the gradual construction (and reconstruction) of a rational social world. It is a familiar feeling in our day-to-day existence to chafe under rules that have not been justified to us and which therefore appear arbitrary. As many children will no doubt testify, “Because I said so!” is not much of a reason. When we discussed authority we noted that even where a political actor claims the right to issue content-independent directives[[194]](#footnote-194) their possession of this right must itself be justified. In the case of a parent and child, a full justification may have to wait until the child reaches maturity. Even so, it would be odd and counter-productive if a parent never attempted to engage with the child about the basis of parental authority.

Out of respect for the individual, liberal theorists insert a publicity condition into the moral justification of states. Waldron (1987, p.149), for example, says the following: “Liberals demand that the social order should in principle be capable of explaining itself at the tribunal of each person’s understanding”. In terms of an intelligible social order there are actually two levels at which the justification of the rules of our political community matters to us.

First, we expect individual elements of the social order to be justified. The BBC (2005; 2014) has reported that the government of North Korea requires its people to follow strict guidelines when it comes to getting their hair cut. It may even be the case that male university students have been officially ordered to have their hair cut in the very same style as their ‘dear’ leader. The reasons that are apparently being offered for this extreme practice are, predictably enough, bad ones.[[195]](#footnote-195) Short hair, for example, does exactly nothing to promote the aims and ideals of socialism. It must be very difficult indeed for members of that society to make sense of their political community and to reconcile their significance as reason-responsive individuals with the totalitarian control exercised by the state.

Being subject to arbitrary power is a lot like playing a game when you don’t understand the rules. It is a very frustrating and alienating experience. It is frustrating because it is extremely difficult to succeed in one’s aims. Indeed, in those pursuits that depend on collective social practices it may be impossible to even determine what would count as success.[[196]](#footnote-196) Setting and pursuing objectives in this kind of scenario is fraught with difficulty and it is maddening to be stymied in this way. It is alienating because you do not feel a part of the game. If you cannot relate to rules or the other players then you cannot understand yourself in terms of your role in the game. Writ large on the scale of a political community, this is a lonely position to occupy and it can lead to self-doubt and paralysing uncertainty.

It can be alienating for citizens even when there is only one particular law, or cluster of laws, that does not stand up to rational scrutiny. Indefensible discrimination based, for example, on race or sexual orientation will lead conscientious people, whether they are directly affected by the law or not,[[197]](#footnote-197) to perceive the state as estranged from reason and a less hospitable place in which to live as a result.

Second, an intelligible social order is, like an intelligible life, concerned with responding to reasons. We want to know that our political authorities understand that the validity of their claim to wield power is contingent on their continuing to serve the public interest by, for instance, solving coordination problems and providing various goods that we need as naturally gregarious animals.[[198]](#footnote-198) If a society is burdened by archaic laws or unacceptable practices then it must be apparent that the state is on the case, and actively engaged in making the world more rational. Subjective acknowledgment that the state is justified and generally doing a good job provides the foundation for citizens to feel at home in their political community. They can then see the police force as protectors, rather than oppressors, and politicians as their agents rather than their masters.

I have argued that one method of responding to reasons that is open to us is to understand ourselves through particular principles that seem to us sufficiently justified.[[199]](#footnote-199) Conceiving of oneself as a good citizen can be viewed as an effective response to reasons. Identification with the state is stable when it grows from identification with principles because then it is buttressed by the sanction of unintelligibility. This is the political equivalent of the state of grace that I described in Section 3.6. Citizens not only identify with the principles that underpin the justification of the state, but also feel that the state is living up to its responsibilities and are happy to support it in whatever way that they can.

Since states serve us partly by contributing to the intelligibility of our individual lives, there is perhaps something helpful that we can say about the balance between personal autonomy and autonomy as identification with the state. The latter stage picks up where personal autonomy leaves off. Individual actions are not always the appropriate way to respond to practical reasons. Sometimes it is through collective action and submission to hierarchical power structures that we can best solve particular problems. It would be a motivation for philosophical anarchism if it was impossible to reconcile our status as agents with subjection to authority.[[200]](#footnote-200) Happily this is not the case. Sometimes a concern for intelligibility mandates individual action, but sometimes it requires supporting collective actions and the actions of others by participating in a legitimate system empowered to deploy coercive force.

We do this, I suggest again, primarily by conceiving of ourselves through the principles that underpin our political authorities. Just as an appropriate response to the considerations that count in favour of helping little old ladies across the road is to think of oneself as the kind of person who helps little old ladies across the road, an appropriate response to the kinds of problems that require state coordination and intervention is to think of oneself as the kind of person who obeys the laws of a just state. In this way we bind ourselves to our political communities. Of course, these principles are efficacious because they have exclusionary force. This means that identifying with a principle restricts the options that one might otherwise have had for responding to relevant practical reasons.

It is a sacrifice to concede some of one’s personal autonomy. People are often tempted, for example, to take the law into their own hands. Locke thought that, under the right circumstances,[[201]](#footnote-201) it is a perfectly appropriate response to an injustice to dole out punishment yourself. However, it is an even better response to institute and uphold a fair and impartial state authority. Supporting a functional criminal justice system entails excluding certain considerations that might weigh in on the side of vigilantism, such as seeing that justice is done swiftly or in accordance with one’s special perspective on the case.

Identifying with principles that mandate obedience to a suitable political authority, therefore, is an elegant and effective way of responding to reasons. The ideal scenario comes about when the justification for the rules and institutions of a state fits with the self-conception of its citizens, and those citizens can see the state responding to reasons well and on their behalf.

It is worth noting that widespread brainwashing might conceivably succeed in producing a subjective sense of well-being if it caused the citizens of a state to believe that their political authorities are just but it would be no good on my view because identification with principles is only a response to reasons if it is itself voluntary. In the next section I will attempt to articulate the foundational interest that almost all of us share in responding to reasons and explain how that takes the form of an intelligible life. I will then return to the subject of voluntariness in Section 5.7.

**5.6 – Autonomy as Reason-Responsiveness**

I have argued that at least one of the reasons why we take personal autonomy to be valuable is because it matters to as agents that we have sufficient control over our lives to respond appropriately to practical reasons. I then showed that identifying with a suitably just state is important because it is a means for individuals to come together and respond collectively to reasons in ways that they could not alone. In this section, my objective is to flesh out the value that we attach to expressing our rational natures by reasoning well and responding appropriately to reasons. Developing this idea will help to elucidate the significance of welding our responses to reasons together to form a settled character and live an intelligible life.

Autonomy is often understood in opposition to heteronomy.[[202]](#footnote-202) We are not autonomous, we might think, when we are driven by desires and appetites as opposed to what our reason tells us. It seems to me that the problem here is not so much that we have needs and wants as embodied creatures with a complex evolutionary past. It is true enough that I can exercise only limited control over my need to eat, for example. But this only undermines my ability to understand myself as a rational agent when my biological imperatives loom too large over my deliberations and obscure other, weightier, considerations for action. There is nothing very worrying about my compulsion to yawn if I see you yawning. If I continue to stand in line for a sandwich while there is a child drowning in the pond beside me, that’s when I have an issue.

This is because it would be unintelligible to acknowledge the force of the reasons that count in favour of rescuing the child and simultaneously queue patiently for my lunch. Such inaction would not be justifiable to anyone else, and on reflection it would be impossible to justify to myself either. It would not even suffice as an explanation, any observer would be forced to make assumptions about either my values or my capacities in order to get any kind of handle on my failure to intervene.

Excepting extreme circumstances of deprivation or stress, we can, as Korsgaard (1996a, p.93) emphasises, always “back up” from our drives and desires and reflect on the question of what we *should* do. What would we make of a creature who possessed this capacity but was never moved to weigh considerations against one another in order to decide how best to act? Very little, I think, and if we do not want to be like that then we have an interest in being responsive to reasons.

But why should we not stop there? In response to Korsgaard, Nagel (1996, p.202) suggests that there is nothing in principle preventing a rational entity from determining itself in fragmentary, disconnected choices.[[203]](#footnote-203) Perhaps I could save the child and act for reasons without committing to a general principle of saving drowning children in ponds. I wish to go further here and recommend a form of life in which one binds oneself to a hierarchy of principles through the cost of unravelling one’s self-conception. How can I support this?

I will appeal to the value of being someone in particular. Nagel is quite right that we can conceive of a rational creature that determines itself in isolated, individual choices. There is, however, no appreciable substance to the personality of such a creature. It cannot be defined by its commitments and therefore does not have a character. It would be remiss of me not to note that labels and definitions can be very dangerous things. That said, it is part and parcel of how we think about ourselves and about others. To be a concrete individual is to identify with principles and thereby grant them exclusionary force, and it is to aspire to structure one’s responses to reasons according to the general considerations that one takes to be important.[[204]](#footnote-204)

To understand oneself as a composer of atonal music, for instance, is to impose a pattern on the considerations that one encounters in the course of composing a piece of music. It is easier and more traditionally pleasing to set a melody in a particular musical key. However, the first-order attractions of the rules of tonal music are sidelined as reasons for a composer who identifies herself as an atonal artist.[[205]](#footnote-205) It is the same for the other identities to which we might subscribe. To conceive of oneself, say, as a member of a family is to impose a family-oriented pattern on the considerations that one encounters. And so on. Responding to reasons does not presuppose a practical identity, but in order to live an intelligible life one must respond to reasons within a framework of exclusionary principles.

Perhaps a more familiar way to approach this point is to emphasise the importance of consistency. We can make sense of a single action that is taken for a reason, without adopting a principle. However, that does not give us all of the intelligibility that we want. Reasons are general in nature. If something counts as a reason in one case then it will count as a reason in a similar case, unless there is a relevant difference between the two. Agents who do not act consistently in similar cases, or who cannot cite a relevant difference, are unintelligible, and this is something we want to avoid. I have argued that, since we care about responding appropriately to reasons, one excellent strategy at our disposal is to organise our self-images around general principles for action.

It should now be apparent why autonomy matters to us primarily in the context of a whole life rather than as a property of individual actions. To evaluate an action as a response to an agent’s reasons we must be able to understand it in the context of her other attempts to respond to reasons and set it against the backdrop of the hierarchy of principles that she has fashioned into a practical identity.

However, we must be careful how we specify the sense in which individuals are committed to living an intelligible life. Colburn (2010, p.26), for example, is eager to avoid “a bias towards the autonomous life being reflective, rather than active.” Although he is concerned with what I have called here personal autonomy, he (2010, p.26) raises the general worry that too much emphasis on deliberation “will tend to crowd out action”. In fact, there is no real tension here. Action is often the appropriate way to acknowledge the weight of reasons and I do not have to hold that we must always be consciously concerned with how well we are responding to reasons in order to live an intelligible life. What matters is that we identify with principles for good reasons and that our actions conform to those principles. This leads me back to the significance of voluntariness.

**5.7 – Voluntariness and Living an Intelligible Life**

In opposition to Rawls and Nagel, I have proposed that a hypothetical consent theory of the justification of political authority can incorporate a substantive concern for voluntariness itself, rather than attempting only to realise some of the values of voluntary participation. On my view of obligation, it is up to each individual to police her own commitments by staking her self-conception on those principles with which she identifies. Since we can identify with principles because of the persuasiveness of the reasons that support them, I contended that we can be bound to obey just states in a way that can meaningfully be described as voluntary. In this section I will elaborate on why this conclusion matters. Specifically, I will argue that a concern for voluntariness itself is worth preserving because it is only when our commitments are voluntary that the binding force of obligation is a response to reasons.

Voluntary commitment to our moral principles matters because deliberate identity formation as a response to reasons is driven by conviction. It is only when I am moved by the acceptability of a principle that it can contribute to an intelligible life by forming part of how I think about myself. This may seem counter-intuitive. After all, innumerable people would define themselves as *against* evils like tyranny, injustice, gross inequality, and so on. But this is an oversimplification.

An Occupy Wall Street activist might rail passionately against the excesses of the one percent and the corresponding hardships experienced by (some of) the other ninety-nine. She would be mistaken, however, to think that the ultimate explanation of her commitment is that she sees that existing social conditions are unacceptable. In fact, if she aspires to hold her views on the basis of reasons then she will justify them by citing values. Unless you have foolishly put all of your money into tulip futures you are unlikely to hold that something is valuable because it is unacceptable for it not to be valuable.

Valuing is a typically voluntary process, as I have defined voluntariness here,[[206]](#footnote-206) because it is grounded in the instrumental or independent appeal of something. Our activist may turn out to be committed because of her conviction that individual human beings matter, and that their value is incompatible with a society that permits massive inequalities in its distribution of resources. Forming an identity is thus rooted in the process of identifying those things that we value. We can then order our responses to reasons according to the principles that seem to us to best express our values. In this way an autonomous person lives what I have called an intelligible life.

And this is why brainwashing would be a terrible mistake, even if it successfully compelled misguided dissidents to support a state that was, otherwise, eminently just. Their actions might now better fit their reasons, but they would not be bound to support the state as a *response* to reasons and would, therefore, not be contributing to their efforts to live an intelligible life. All of this is not to deny that there are important procedural conditions that must obtain if an individual is to have the capacity for autonomy, but in the next section I will argue that an account of autonomy should be thought of as substantive in one vital respect: an autonomous individual must actually value the project of living an intelligible life.

**5.8 – Procedural and Substantive Conceptions of Autonomy**

Before I conclude this chapter I want to spend a little time teasing out one implication of my account of an autonomous life as one in which an individual binds herself to a hierarchy of principles as a means of responding to reasons. In this section I will argue that theories of autonomy that are purely procedural are unable to capture how a commitment really belongs to an agent because they fail to consider how agents are bound to their commitments. My suggestion is that autonomous individuals must actually care about living an intelligible life and that this ties them to the principles that they incorporate into their identities.

Some basic cognitive capacities, and the effective ability to exercise them, must be present for an entity to be capable of responding to reasons. Some theorists think that this is ultimately all that is required for autonomy. They offer procedural accounts, where autonomy is a function of some group of capacities that an agent can exercise. When these capacities are properly protected the integrity of an agent is guaranteed, and we are thus entitled to say that she[[207]](#footnote-207) is autonomous. There are many such accounts. Frankfurt (1971; 1988), for instance, focuses on wholehearted identification with one’s second-order desires, whereas Christman (1991a; 1991b) looks at personal history to ensure that an individual’s self-reflection has not been inhibited, and Marilyn Friedman (2000, p.36) prefers reflective endorsement so that she can “encompass emotional, as well as strictly rational or narrowly cognitive dimensions of personal processes”.[[208]](#footnote-208)

My account is a substantive one. There is some specific content to which I think individuals must be committed in order to live an intelligible life that is meaningfully their own. They must actually care about living an intelligible life. It is only if this matters to them that they can deploy the mechanism of binding force as I have explicated it here. If they recoil from the possibility of unravelling their identities and if they attend to practical reasons, partly at least,[[209]](#footnote-209) in light of the overarching goal of becoming a particular person, then they are in a position to tie their self-conceptions to a hierarchy of principles. By binding themselves in this way they create concrete identities and engage with normative considerations as unique individuals.

Let’s say that I form a second-order desire to take up smoking. Smoking is cool, and all the other kids will think that I’m cool if they see me casually puffing away. I reflectively endorse this desire. When I think about it, smoking still seems like a good idea and the outcome a desirable one. However, smoking makes me cough and I find the taste unpleasant. My desire might be strong enough to carry me through such initial obstacles. Then again, it might not. If, in the second case, I do not feel bound in some tangible way to persevere and push through these obstacles then I do not see how my second-order desire is helpfully explained as ‘mine’.

Procedural conceptions fall down by neglecting the significance of bindingness for integrity. I may reflectively endorse desires, beliefs, actions, or principles without plausibly making them my own. They are only a part of me if I am committed to them, and I am only committed when I am bound to them. Even if you reject my account of obligation, it should hopefully be apparent that bindingness is relevant to developing an account of what it means to live an autonomous life. Procedural accounts of autonomy are deficient in this regard since they neglect to consider the significance of bindingness generally.

**5.9 – Conclusion**

In this chapter I have argued for a conception of autonomy as living an intelligible life by responding to reasons within a framework of exclusionary principles. This conception ties together three common senses of autonomy that I have labelled personal autonomy, identification with the state, and reason-responsiveness. My contention was that the significance of reason-responsiveness was the key to understanding the underlying unity of these conceptions. They are all constitutive elements of an intelligible life.

In order to motivate the value of such a life I appealed to the idea of forming an identity and becoming someone in particular. I explained the significance of voluntariness for our political obligations in terms of the way that we value things and the importance of responding to reasons by instantiating those values in the principles that constitute our identities.

Finally, I distinguished my view from procedural conceptions of autonomy on the basis of the substantive content to which I think individuals must be committed if they are to construct their own practical identities and respond to reason in an ordered fashion. It is only if they already care about living this sort of life that we are entitled to say that they are bound to their commitments.

In the next chapter I will return to the methodology of the social contract and argue that we would do well to invert it so that we can develop a picture of the conscientious citizen who relates to her political responsibilities in the right way. I will apply the conception of autonomy outlined in this chapter by emphasising the significance of identity and identity formation as a response to reasons for establishing the stability of a conception of justice and for securing the possibility of progress towards our ideals.

# 6

# *Inverting the Contract*

**6.1 – Introduction**

In this final chapter I will apply some of the ideas that I laid out in earlier chapters in order to demonstrate their usefulness. My overarching aim will be to motivate a new strategy to complement the classic social contract model. Traditionally, contract theorists begin with some definite conception of the person and then design principles to which it would be sensible for such a being to assent. My proposal here is that we should also focus on developing the practical identity that a citizen must have if she is to bind herself to a political community. I will refer to this identity as a contracting identity, as outlined in Section 4.3, and I will call the approach of constructing a practical identity from the building blocks provided by a standard theory ‘inverting the contract’ because it asks what a person will have to be like in order to relate to the conclusions of a standard contract argument in the right way to generate the characteristic force of obligation.

The basic idea is that there is only so much to be gained by making the chains that citizens are to wear more comfortable.[[210]](#footnote-210) There is, however, scope to work from the opposite end by considering the self-conception that citizens will require if they are to reconcile themselves to the constraints that come with living alongside others in a coercive state. Although it must remain up to individuals to embrace an identity that is compatible with their political responsibilities, contract arguments are uniquely equipped to deploy the characteristic normative force of obligation because the prospective parties to the agreement model a commitment to basic principles with which most of us already identify. As the theory unfolds and new principles are derived, we can imagine the bare identities of the artificial parties being fleshed out. This provides a template through which we can understand our own self-conceptions. We will then be in a position to give a complete picture of political obligation: the just state embraced by a conscientious citizen.

Aside from the satisfaction of a more complete political philosophy, there are some smaller problems that become tractable when we exhume the link between hypothetical consent and obligation as I have suggested. I will attempt to demonstrate the merits of my proposal by explaining how inverting the contract can help with these issues. If I am successful in this enterprise then we will have grounds for applying the strategy more widely.

First, I will argue that Rawls’s concerns about the stability of a conception of justice must be addressed by incorporating an identity-based account of the binding force of obligation into the overall justification of a contract theory. Stability can only be achieved when citizens incorporate the foundational principles and specific demands of a conception of justice into their self-conceptions and thereby secure for those principles a regulative role in their deliberations underpinned by the sanction of unintelligibility.[[211]](#footnote-211)

This analysis also provides a more convincing explication of how citizens can hold different comprehensive doctrines and yet come together to form an overlapping consensus. On my account, the principles established by a contract theory are to play an exclusionary role that dictates the conditions under which citizens licence themselves to rely on their non-political values. Although Rawls would likely disagree with my interpretation, I shall defend the significance of an overlapping consensus as the first stage in inverting the contract since it facilitates prospective citizens from a multitude of comprehensive doctrines in affirming their commitment to the political principles of justice.

Changing tack, I will argue that inverting the contract approach can give a new impetus to ideal theory, which has come under attack for, among other things,[[212]](#footnote-212) failing to provide transitional guidance towards a more just society in our non-ideal circumstances. When we consider the interdependence of how citizens think of themselves and the conditions under which the demands of justice can be realised, we can see how important it is to promote a contracting identity. With the right identity, citizens will be less likely to become complacent about the *status quo* or to settle on short-term benefits that imperil the continued pursuit of perfect justice. This is because, for them, progress towards the ideal remains obligatory. On this view, ideal theory does not need to generate concrete policy suggestions for non-ideal circumstances in order to be useful. Instead, it serves to dictate central elements of the identities we should strive to express in more or less hostile conditions.

I will begin by examining the role of the conception of the person in classic contract theory and contrasting it with the approach pioneered by Plato. In Section 6.3 I will consider more precisely how the binding force of obligation can be utilised by contract theorists, distinguishing between obligations and obligatory ends. Section 6.4 will address the issue of stability, while the following section will deal with the notion of an overlapping consensus. Before concluding I will turn, in Section 6.6, to address the ideal vs. non-ideal theory debate.

**6.2 – Rational Reconstruction of the Individual**

For Rousseau, the question that prompts the social contract tradition is how to make coercive power legitimate. It is natural to see the task here as figuring out the moral principles that should underpin a state and then designing a state to meet that standard. However, I want to propose a complementary approach. In offering a view about the moral foundations of the coercive state, a social contract theory has to do something else; it has to decide on a view of human beings. What matters to them and how do they work? The trick is not to envisage a society in which there are no chains, nor is it just to justify the chains that do exist. We must show how those chains can be borne.

Jean Hampton understands the classic social contract theories as attempts to provide a rational foundation for the state, which requires one if it is to be able to justify its existence to its citizens who must continue to support and maintain it. She holds that “the social contract argument is designed as a certain kind of *rational reconstruction* of the state… the social contract theorist figuratively takes the state apart and (rationally) puts it back together again in order to explain the structure of any *existing* state and to determine whether or not it is justified” (Hampton 1986, p.269, original emphasis). Hampton comes close, but ultimately fails to see that a social contract theory is not limited to the rational reconstruction of state, but can also attempt the rational reconstruction of the individual[[213]](#footnote-213) who is to be a citizen.[[214]](#footnote-214)

This is not, by any means, a novel idea. Plato, I believe, has something very similar in mind in *The Republic*, where he thinks that he can draw a sensible analogy between justice for the individual and justice for the state,[[215]](#footnote-215) and even goes so far as to sketch the kinds of characters that mirror different forms of political organisation.[[216]](#footnote-216) His thought, if I interpret him correctly, revolves around the idea that if we understand ourselves in different ways – as philosophers, or warriors, or businessmen, or atomised individuals, or desire-gratifiers – then these different worldviews will incline us towards societies that express and promote the corresponding values.[[217]](#footnote-217) The degeneration of the state and the citizen go hand in hand because in the absence of pressure to conform to the correct set of values individuals will come under pressure to realise others,[[218]](#footnote-218) and they will organise their plans around whichever one comes to dominate.[[219]](#footnote-219) His point is that the benefits that a just state can achieve for its people is proportional to their commitment to justice, and their commitment naturally wanes if it is not reflected back to them in their institutions and made a concrete part of their everyday lives.

In more modern parlance, we may rephrase this as the pursuit of stability in two domains: the state and the individual. A stable political community is one that has enduring institutions and shared values. A stable individual maintains a consistent self-image over time and regulates her actions according to a hierarchy of principles. This allows her to live an intelligible life.[[220]](#footnote-220) These go together because almost all of us need external assistance to piece together a coherent identity in the first place and, with the aid of various social pressures, to persist with that identity when beset by temptation and misfortune. A stable state, on the other hand, relies upon the goodwill and reflective support of its citizens to function efficiently and to successfully assert its authority in the event that its legitimacy is ever questioned or undermined.[[221]](#footnote-221)

Rawls (1999, p.398) says that on the theoretical level, “[o]ne conception of justice is more stable than another if the sense of justice that it tends to generate is stronger and more likely to override disruptive inclinations and if the institutions it allows foster weaker impulses and temptations to act unjustly”. This is to say that a conception of justice is not fit for purpose if it will not, over time, generate support for itself, which is why he (1999, p.436) later says that: “The most stable conception of justice, therefore, is presumably one that is perspicuous to our reason, congruent with our good, and rooted not in abnegation but affirmation of the self”. In Section 6.4 I will discuss stability in more detail.

Of course, Plato is not primarily concerned with the concept of political obligation but his holistic approach is instructive nonetheless. In tandem with a justificatory approach we need an account of what it will be for an individual to internalise those principles that emerge and to internalise them is, I have argued, the same as to make them a part of one’s practical identity. So, like Plato, the social contract theorist needs an account of the practical identity that goes alongside the principles that she recommends.

Hampton (1986, p.270) rejects this parallel because she believes that contract theorists all share the same method of beginning with a more or less determinate conception of the person and working out what can be justified to her: “In contrast to philosophers such as Socrates and Aristotle, the social contract theorist believes it makes sense to think of individuals as conceptually prior to the political systems in which they live, and to this extent the contractarian is an individualist”.[[222]](#footnote-222) While correct, Hampton’s emphasis here is, I think, misguided.

Contract theorists idealise the participants who are to agree, on our behalf, to the hypothetical agreement that they devise for two reasons. First, the parties should be procedurally better reasoners than we are since we do not want them to be misled by false beliefs and invalid arguments, or get sidetracked by the pressures and vagaries of everyday life. Second, contract theory is not an exercise in deliberative democracy.[[223]](#footnote-223) In order to guarantee a result it makes sense to limit the plurality of perspectives and commitments of the parties at the outset, up to and including the position Rawls takes where they can be reduced to a single viewpoint.

In so far as these features must be decided upon before any theorist begins in earnest, Hampton is correct. However, the level of abstraction with which one approaches the second consideration is key to determining how best to relate the representative parties in the initial situation to real people. The more abstract the conception, the more scope there is to build the framework of a practical identity from the commitments that underpin the agreement and the principles that follow from it. It is also easier to show the space that would remain for the concrete individual to pursue additional identities and goals within the framework of her contracting identity. Reasoning from whatever starting point one takes, a hypothetical consent theorist aims to establish what further principles someone who is answerable to the idealised procedural standards and starting commitments would accept.

It is striking that Rawls opts to reconstruct this process in such a clearly exclusionary form.[[224]](#footnote-224) As the theory unfolds, investment in the initial principles and considered judgments is (all going well) shown to have profound consequences in the shape of further ends and principles to which the parties must, on pain of unintelligibility, also be committed. Starting with a more or less stripped down identity, the contract theorist gradually fills in the detail, sometimes by appealing to additional values or considerations as the discussion becomes more concrete. What Hampton misses is that the conception of the person evolves as the theory develops.

So, just as the state is taken apart and put back together again, a template for the prospective citizen can be built up from the basic principles that the theory takes as its starting point, be they respect for persons, enlightened self-interest, or anything else. This means that the parties and the hierarchy of their eventual commitments can serve as a model[[225]](#footnote-225) that those of us who subscribe to the starting principles have reason to consider very carefully as an ideal for approximation.

**6.3 – Obligations and Obligatory Ends**

The representative parties embody the conception of the person at the heart of a contract theory. What guidance can they offer us? In order to answer this question we need to start by distinguishing two ways in which the binding force of obligation can be experienced. It can constrain us with regard to actions, rendering them either unacceptable or required, or it can constrain us with regard to ends, making them either unacceptable or required. This corresponds roughly but not exactly to the Kantian categories of perfect and imperfect obligations.[[226]](#footnote-226) An example will help.

Let’s consider again Rawls’s natural duty of justice.[[227]](#footnote-227) In his version of the initial choice situation the parties would agree that they must always comply with the directives of just states as they exist and apply to them. If you agree with the way that Rawls sets up the original position and that the veil of ignorance excludes considerations that should not count in determining how one ought to relate to a just state, then you either have to accept the duty that he posits or face the prospect of damaging your self-understanding. The first option has the following consequence: you have to do what a just state acting within its remit tells you to do and you must refrain from violating its laws. Certain actions are mandatory for you and others are proscribed. Most of the time you can satisfy this duty simply by abstaining from wrongdoing and in the instances when the state demands that you perform an action the demand is not open-ended. You must, say, pay some percentage of your income in tax by a set date, or you must send your children to school until they are 18. You are either in compliance with the duty or you are not.[[228]](#footnote-228)

There is another element to the duty to obey just states, which is that we must support them and “further just arrangements not yet established, at least when this can be done without too much cost to ourselves” (Rawls 1999, p.99). This is less onerous insofar as it leaves us some discretion. I can direct my efforts at nurturing State A rather than State B, or I can dedicate myself to putting in a little extra time supporting justice every other Friday and pursue my other goals the rest of the time. However, it is open-ended. We could always do more to support just states that already exist and to nourish just institutions in states that have yet to become fully, or even tolerably, just.

It may seem that such ends cannot be truly obligatory on my scheme since the discretion that is built into them appears to entail that they have no exclusionary power, but this is not the case. To be committed to an end one must be prepared to exclude other considerations when the circumstances demand it. This is because there are times when anything other than an action that promotes the end is incompatible with conceiving of oneself as having that end. A British citizen who declined to pitch in during World War II in at least some way could hardly claim to genuinely hold as an end the safety and prosperity of a liberal democratic Britain. What Kantians sometimes call the ‘latitude’ of the commitment has receded.[[229]](#footnote-229) Although a wartime citizen might be able to exercise her judgment in deciding how best to help the war effort, she cannot decide that she needs to catch up on her reading right now but will definitely help in the next war instead.

Another difference between obligations and obligatory ends is that the reasons that support the latter give us cause to look for more opportunities to promote the end whereas there is no reason to look for more laws to comply with. One of the ways in which the state helps us to respond to our individual reasons as a collective is that it can make our obligatory ends more determinate.[[230]](#footnote-230)

One of Kant’s examples of an obligatory end is the duty of assistance.[[231]](#footnote-231) Let’s agree that we have sufficient reason to commit to a principle of helping people in need of assistance. Now, it goes some way to realising that commitment if a portion of my tax contribution is pooled with portions of everyone else’s tax contributions and used to fund ambulance and emergency care services. Indeed, we would have to question very seriously the commitment of anyone who resented paying towards the service or who would opt out if offered the chance.

Building from some basic principles the representative parties spell out how an original commitment can be extended around further principles. For an individual with a compatible practical identity this process can generate the binding force to secure both obligations and obligatory ends because it implicates her intelligibility to herself. However, since the parties are, by design, a much simplified rendering of those features of ourselves that are most relevant, it is critically important to draw a line from the representative parties to flesh and blood human beings who hold a multitude of ends and sustain many nested identities.[[232]](#footnote-232)

The later Rawls was, of course, very careful to point out that the parties in his original position are not abstract versions of ourselves.[[233]](#footnote-233) They are merely our representatives. This is meant to insulate him from the general worry about hypothetical consent, as well as any suspicions that he is advocating a ‘metaphysical’ conception of the person. This suggests rather strongly that he would resist any attempt to bulk up the significance of the parties and their hypothetical consent.

In spite of his hesitancy to advance a metaphysical view or a comprehensive doctrine, and although nothing much is supposed to hang on it, he takes great pains to demonstrate how life is likely go better for individuals who assimilate the principles of justice that he is offering. He refers to this idea as the congruence of the right and the good, or as goodness as rationality: “In some circumstances at least, for example those of a society well-ordered or in a state of near justice, it turns out, I believe, that being a good person is indeed a good” (Rawls 1999, p.349). Congruence is important for the sake of stability.[[234]](#footnote-234) In the next sections I will show that Rawls’s approach to stability only works if we if we recognise the importance of facilitating prospective citizens in taking up the contracting identity modelled by the representative parties. More specifically, it is only by conceiving of it in this way that we can make any sense of his pursuit of an overlapping consensus in *Political Liberalism*.

**6.4 – Stability**

Stability for a conception of justice is achieved by producing a strong sense of justice in citizens and permitting only institutions that promote conformity with the underlying principles of justice.[[235]](#footnote-235) For Rawls (1999, p.41), a person has a sense of justice if she possesses the intellectual capacity to judge things to be just or unjust on the basis of reasons and is, crucially, motivated to act in accordance with these judgments.[[236]](#footnote-236) However, we might well ask how an ideal theory can get a hold of a person so that they become disposed to act on particular principles of justice. Rawls’s original answer was that it is ultimately in their self-interest to develop a sense of justice. But there is a gap here. Rawls needs to get from self-interest lending weight to considerations of justice in our deliberations to a *regulative* sense of justice. More particularly, I think he is mistaken in painting the issue, as he sometimes does,[[237]](#footnote-237) as one about the balance of our self-interested reasons. Stability, I submit, is not satisfactorily achieved by uncovering weighty enough reasons to offset citizens’ temptations to depart from the principles of justice. Citizens must be meaningfully bound to acknowledge reasons of justice. Stability requires a stick as well as a carrot.

In this section, and the following section, I will argue that for stability to exist citizens must feel bound to adhere to the particular principles of justice that a theory specifies. I will deploy this analysis to suggest a way to understand and flesh out Rawls’s puzzling claim that individuals from radically different comprehensive doctrines can, from within those doctrines, form an overlapping consensus and endorse limits to their pursuit and expression of those worldviews.[[238]](#footnote-238) However, we should not stop at establishing the possibility that a sense of justice could be consistent with a wide range of comprehensive doctrines. The next step is to get clear on the details of the regulative role that a contracting identity must play in order to match the citizen’s overall practical identity to the state she is to support and uphold.

For Rawls (1999, p.6), a stable scheme of cooperation “must be more or less regularly complied with and its basic rules willingly acted upon; and when infractions occur, stabilizing forces should exist that prevent further violations and tend to restore the arrangement”. All this is to say that any set of rules for a mutually beneficial association will not get very far unless the participants are moved not only to adhere to it, but also to promote and preserve it when breaches occur.[[239]](#footnote-239) But where is this motivation to come from?

It is important to stress that Rawls is not satisfied by a mere *modus vivendi*. As Brian Barry (1995, p.882) points out, just institutions cannot be maintained if enough people do not support them willingly. Fear might well hold some systems together, but they will not be just systems. Nor does Rawls want to depend on any assumptions about the intrinsic motivational force of morality. He (1999, p.499) shies away from the doctrine of the pure conscientious act. What’s left is to prove that it is in the self-interest of individuals to nurture their sense of justice.

He (1999, p.359) argues that a person’s good is the pursuit of a rational plan of life since “a rational plan establishes the basic point of view from which all judgments of value relating to a particular person are to be made and rendered consistent”. He (1999, p.361) goes on to elaborate that: “A plan then is made up of subplans suitably arranged in a hierarchy, the broad features of the plan allowing for the more permanent aims and interests that complement one another”. So, the question of congruence then becomes a question of whether it is good for any individual to constrain their rational plans by the principles of justice.[[240]](#footnote-240)

There are a number of important considerations here. For one thing, it is no easy task to decide what one’s plan of life should be and the “priority of right and justice securely constrains these deliberations so that they become more manageable” (Rawls 1999, p.494). For another, Rawls thinks it is a precondition on engaging in a social union of social unions in such a way as to get the most out of it. His Aristotelian principle states that we seek to deploy as many of our skills and capacities as possible and the zenith of challenging activity is communal action, which is only possible if people are prepared to treat one another fairly.[[241]](#footnote-241) Finally, it facilitates the building and maintenance of all sorts of meaningful relationships with others that would not be possible without a deep and public commitment to the principles of right.[[242]](#footnote-242)

The general strategy that Rawls adopts is to add more weight to one side of our “balance of motives” (1999, p.501), but here’s the rub: he doesn’t just need to make justice a bit more attractive; he needs to transform the way that the citizen relates to reasons of justice. I say this for two reasons. The first has to do with the factors that lead to instability. He (1999, pp.295-296) discusses two such factors: the attraction of free-riding and the possibility that others are not doing their fair share. To those I would add two more: our own good is only one consideration among others and human beings have gotten very good at selectively attending to particular considerations. That having a sense of justice is in my interest[[243]](#footnote-243) does not obviously outweigh other values that I might hold, nor does it prevent me from zeroing in on some consideration that attracts me even though I know deep down that it is outweighed by others.[[244]](#footnote-244) Rebalancing the scales does not secure the influence of reasons of justice and will likely fail to make much of a difference in the kinds of cases when we are tempted to make exceptions to rules that we acknowledge.

The second reason is that he recognises it himself: “Thus, what is to be established is that it is rational (as defined by the thin theory of the good) for those in a well-ordered society to affirm their sense of justice as *regulative* of their plan of life” (Rawls p.497, my emphasis). And again: “The problem is whether the *regulative* desire to adopt the standpoint of justice belongs to a person’s own good when viewed in the light of the thin theory with no restrictions on information” (Rawls 1999, p.497, my emphasis). I have already explained how a principle or a reason plays a regulative role[[245]](#footnote-245) – it excludes some considerations from counting in our deliberations – and I have argued that exclusionary force must be backed up by a sanction. What stability requires, therefore, is that citizens are bound by reasons of justice through their own self-conceptions.

This fits surprisingly neatly with how Rawls depicts our moral psychology in Chapter VIII of *A Theory of Justice* and how he (1999, p.428) thinks he can appeal to someone who already has the requisite self-image: “Now, the fact that one who lacks a sense of justice, and thereby a liability to guilt, lacks certain fundamental attitudes and capacities is not to be taken as a reason for acting as justice dictates. But it has this significance: by understanding what it would be like not to have a sense of justice – that it would be to lack part of our humanity too – we are led to accept our having this sentiment”.[[246]](#footnote-246)

We can see the usefulness of thinking about stability in terms of a contracting identity if we consider a related point. Rawls neglects to pursue the question of how a sense of justice would work in practice. It is clear enough that he doesn’t expect citizens to reflect on its long-term value in each and every circumstance in which they are attracted by the prospect of reneging on their commitments. So a sense of justice must guide our deliberations in a different way. The most straightforward explication is that you have a sense of justice when you *identify* with the demands of justice. As such, it regulates your conduct in the same way as any other part of your identity.

Rawls thought that stability was key and described *Political Liberalism* as an attempt to remedy the deficiencies of his earlier account.[[247]](#footnote-247) Indeed, Barry (1995) takes him to task for misunderstanding his own theory. In the next section I will argue that Rawls was concerned with how a prospective citizen could see herself as bound to the state as part of her overall system of values. Although he does not go as far as I suggest, and would almost certainly disagree with the method I propose for taking contract theory forward, I think that he recognised the significance of citizens’ sharing of a contracting identity that takes precedence over their other identities, which might otherwise clash.

**6.5 – The Significance of an Overlapping Consensus**

The major shift that Rawls makes, as I see it, is away from trying to prove that it is rational for most people to develop a sense of justice. As Barry repeatedly points out, Rawls is only speaking to those of us who are reasonable[[248]](#footnote-248) to begin with and so already accord some weight to the principle of respecting persons as free and equal. However, the desire to cooperate with others on fair terms is only a very small part of any particular individual’s comprehensive doctrine. Problems arise depending on the place it occupies within that larger scheme. It is not much use to a political community if someone is disposed to treat others fairly but only if his priest says it’s OK, or if a quick consultation of the hedonic calculus gives the thumbs up.

In fact, a well-ordered society will need its citizens to subordinate large and important elements of their worldviews to the principles of political justice under some circumstances. This will mean, for example, that a staunch pro-life advocate will acknowledge a woman’s civil right to choose if abortion is legal and the decision-making structures of the state are themselves just. He will no doubt continue to think that anyone who terminates a foetus commits a grave wrong and that he has a moral responsibility to campaign against it. Even so, these considerations will have to be excluded as genuine reasons for acting in a way that is incompatible with the rights of patients and doctors as laid down by the state. On my interpretation, it matters that Rawls can demonstrate the possibility of an overlapping consensus because by doing so he shows that it is possible to subscribe to a reasonable comprehensive doctrine and yet regulate its values in such a way that an adherent can also be a citizen of a politically liberal state. Let me explain.

Sooner or later, extremist religious sects become fed up with the intransigence of mainstream followers. It is, sadly, predictable that they will eventually explain the failure to rise up and embrace their doctrines as evidence that the masses have been corrupted and are no longer really, for a topical example, Muslims. On the extremist worldview, you simply cannot both be a true believer and permit a society in which the most inflexible interpretation of the demands of the relevant religious text are not rigorously imposed. Rawls would characterise such people as unreasonable and he is quite right that it is all but impossible to engage with them for the purposes of building a society.

Moderates are, thankfully, a very different story. Most religious believers think that can still call yourself a Catholic, for instance, even if you accept that the state can grant marriage rights to homosexuals, and this is because respect for persons is already part of the mainstream Catholic worldview. My thought here is that what Rawls is looking for from an overlapping consensus is the possibility that one can identify with a particular comprehensive doctrine while assigning a regulative role to the principles that we must all share if we are to count as reasonable. It is not automatically inconsistent with a reasonable comprehensive doctrine to have a sense of justice. Indeed, Rawls needs it to be the case that it can be partly expressive of that general philosophy.[[249]](#footnote-249)

This interpretation is supported by Rawls’s skittishness about the good in his introduction to *Political Liberalism* and the way he sets up the subject of the book. He is very careful to stress the disparity of reasonable comprehensive doctrines and to emphasise that his project will only succeed if all reasonable people can acknowledge the benefits of a political community and affirm the public perspective from *within* their own general worldview.[[250]](#footnote-250)

If citizens can endorse the political principles from within their comprehensive doctrines then it is plausible to think that those principles could govern others within that worldview. So, a Catholic might think that abortion is wrong but also hold that the ways you can oppose it are severely restricted by the requirement of treating other people as independently valuable agents. Respect for persons is already embedded in reasonable comprehensive doctrines so there is no real tension when adherents are required to keep their private views to themselves.

If we are to understand this in an exclusionary way, though, we need to be clear about the role that principles of justice are intended to play for the individuals who are to be subject to them. A contracting identity provides an internal framework that encompasses and organises the more concrete identities that can be regulated by, and thereby made consistent with, the principles arising from the contract process.

The further question, then, is whether it is enough to show simply that one *can* have a regulative contracting identity as part of broader, but of course still reasonable, comprehensive doctrine. One might think that it is. If successful it shows that a politically liberal society can be stable. Actual stability might just be something that has to be generated by the interactions between citizens and between citizens and the institutions of the state within a just basic structure.

Rawls (2005, p.141) eventually boils stability down to two questions: whether people who grow up under just institutions will naturally acquire a sense of justice and whether an overlapping consensus is possible. The first concerns the development of children born into a just society, while the second concerns the legitimacy of imposing the principles of justice on a radically pluralist group of people. But there is a third question. Does his theory have the motivational resources to tie existing individuals to the principles of justice? It is meant to be a theory of justice for us, and, as such, it must speak to us.

This is why we need a sharper focus on the kinds of considerations that are excluded by a contracting identity and when citizens must ignore them. The existence of a large proportion of reasonable people on its own is simply not enough to guarantee the stability of a politically liberal society. For that we need to be able to say something about how a sense of justice can play a regulative role, i.e. how the binding force of obligation works, and to spell out as best we can the priority that it takes over other elements in a citizen’s worldview which she may care about more intensely or to which she might attach more ultimate significance.

In fact, it is only with this background in place that Rawls’s original strategy in Chapter IX of *A Theory of Justice* can make any headway. In order to show that being a citizen of a politically liberal state is something that it is good to be we need to be able to match, as Plato attempted to do, the identity of the individual with the ethos and institutions of the state. Sufficient harmony here is what ultimately allows the state to serve the interest of its citizens by extending their practical identities and thereby helping them to collectively respond to reasons. This means that the stability of the theory and the stability of the matching identity are interdependent and a concern for both should be incorporated into the justificatory social contract itself.

Now, the ability of contract theories to guide us *at all* in the pursuit of a more just world has been the focus of sustained attack in recent years and I will turn to discuss the controversial debate between ideal and non-ideal theories in the next section.

**6.6 – The Role of Ideals and Ideal Theory**

Social contract theory is a form of ideal theory, and ideal theory has come in for a lot of criticism in recent years although a lot of the back and forth has been at cross-purposes. Laura Valentini (2012) identifies three distinct sets of issues that get wrapped up in the ideal vs. non-ideal debate and it is worth pausing a moment to pick them out. One family of concerns cluster around Rawls’s distinction between full and partial compliance theory.[[251]](#footnote-251) Do we have, for instance, different duties depending on whether everyone else can be assumed to discharge their own responsibilities?[[252]](#footnote-252) Another group of worries has to do with the feasibility constraints that are appropriate and useful for theorising in political philosophy.[[253]](#footnote-253) What kinds of things do we need to factor into our thinking if we are to come up with helpfully action-guiding principles for the often messy and imperfect ‘real’ world? And the third set revolve around the question of whether we need an ideal goal or standard in order to make progress in the pursuit of justice.[[254]](#footnote-254) In this section I want to make a relatively small point to further illustrate the merit of my proposal that a contract argument should be inverted in order to develop a contracting identity for prospective citizens. I will link the second and third set of issues by showing how establishing an end-state can impact on the feasibility conditions that non-ideal theory aims to accommodate.

It sets constraints on the kinds of identities that non-ideal compromises can foster in citizens. This could be understood either as a strong or a weak claim. On the strong version, any institutional or policy concessions to imperfect circumstances must not hamper states in pursuing their responsibility to persuade citizens to identify with the moral justification of rightful political authority and a just basic structure.[[255]](#footnote-255) The weak version simply insists that such concessions do not encourage citizens to form self-conceptions that are incompatible with coming to feel bound to realise the demands of ideal justice. I will defend the stronger claim here.

Simmons (2010) has argued, persuasively I think, that ideal theory must take priority over non-ideal theory if we are to be able to tell what non-ideal strategies and policies are morally permissible and so that we can determine which injustices are more grievous. Ideal theory sets the normative backdrop against which non-ideal debates can meaningfully play out. One of the most significant ways that it does this is by setting our obligations and what I have called our obligatory ends.[[256]](#footnote-256)

No matter how comprehensive and successful an ideal theory might be, there will always be a vital role for a normative approach to the problems of noncompliance by less than perfect people in the often messy ‘real’ world.[[257]](#footnote-257) Non-ideal theory may well call upon us to make compromises that run against the ideal principles to which we are committed. Yet even here it matters how people think of themselves. Farrelly (2007) attacks Rawls on the basis that ideal theories are irrelevant if they can only be constructed by ignoring or abstracting away from the conditions under which any solution would actually have to be realised.[[258]](#footnote-258) The concern is that they offer no guidance at all in the non-ideal world and are thus completely unfeasible. But as Simmons suggests: “suppose that we could structure our basic institutions so that the persons living under them would feel valued and respected as individuals and would in consequence be committed to the principles that guide their shared activities.... If such changes are realistic, then many of the facts about the nonideal world we currently inhabit could cease to *be* facts” (Simmons 2010, p.32, original emphasis). I believe that we are now in a position to see an important limitation of extreme forms of non-ideal theory.

If you rigidly take people as they are then you bracket the possibility that they might change. Of course, changing somebody’s behaviour does not automatically change their identity, but if you can extend an identity in the way that I have been suggesting then you are very likely to change behaviour. The only concession that ideal theorists need really make is to show how their principles would fit together as an identity that can be incorporated into an individual’s self-conception and to sketch how that identity could regulate other elements of the self.

An ideal theory of justice presents a vision of what a perfectly just society will look like, at least in its outline. Competing theories vary the weight of different values and apply more or less stringent feasibility conditions, but they share an aspirational focus on how human beings might relate to one another if they could be relied upon to meet those moral obligations that an ideal theory either sets or relies upon. For this to be the case, I have argued, individuals must conceive of themselves in the appropriate way.

Identity and identity formation are, therefore, key to realising the ideal. However, it is also of huge significance as we look to make progress towards that ideal. If we agree with Simmons that ideal theory sets a goal that we have sufficient reason to pursue, even at the expense of short-term gains, then we have sufficient reason to give it a regulative role over our behaviour here and now by promoting corresponding contracting identities. This is why I prefer the stronger to the weaker claim. Only when citizens have the right self-conception will they feel bound to strive for the ideal and to oppose settling for more easily achievable targets that would, however, still improve on the *status quo*.

In fact, non-ideal theorists need to pay closer attention to the self-conceptions of individuals and the ways in which their practical identities can be developed by the process of considering the demands of justice even before coming to the question of how to implement them. Tethering progressive policies and more concrete principles to regulative elements of a person’s practical identity would allow a non-ideal theorist to employ the binding force of obligation and thereby provide a solid foundation for lasting change. A world that is pervasively unjust will not be a world in which someone who identifies with an ideal theory of justice can satisfactorily express his identity and so something will have to change. Either he will have to find some way to modify his hierarchy of commitments or he will have to go out and try to change the world. Sometimes he will change the world.

This is the power of ideals, they can inspire us to transform the world rather than accommodating ourselves to features that we find repugnant, indeed unacceptable. Obviously this strategy is going to be constrained by what is physically possible and will also run up against the limits of what is pragmatically achievable given social conditions and the other values and projects that individuals hold dear. That said, the idea is that the binding force of obligation can be harnessed to improve matters, and thereby approximate the conditions under which the duties of ideal theory can be discharged.

On this line, critics of ideal theory miss their mark. It is not so important to be able to explain how ideal principles can provide a specific remedy for each and every case of injustice. Rather an ideal theory guides by contributing the impetus to change our circumstances and thereby progress towards a social order that is not characterised by a mismatch between the identities that we aspire to realise in our everyday lives and the circumstances in which we find ourselves.

**6.7 – Conclusion**

The focus of this chapter has been to argue in favour of inverting the traditional social contract model to develop an outline of the ideal citizen. Our discussion of obligation has shown that its binding force takes the form of a sanction that individuals must ultimately exact from themselves. They do this by electing to understand themselves under particular practical identities which are constituted by a commitment to a hierarchy of principles. I have proposed that a social contract theory could be understood in terms of the development of just such an identity – a contracting identity. In this chapter I have explained how a contract argument could ‘rationally reconstruct’ its prospective citizens and how representative parties can model our established commitments and the further principles that follow from them that we must acknowledge on pain of unintelligibility.

I then proceeded to address two areas in which contract theorists would be advised to view contract theory in this light: the stability of a conception of justice and the debate between ideal and non-ideal theorists. First, in so far as we are concerned about how prospective citizens can resist the temptation to disregard reasons of justice for the sake of their individual desires and values we must be concerned with how particular principles of justice can have exclusionary force for an individual. However, it is not enough to show, as Rawls attempts to do, that having a sense of justice could be compatible with reasonable comprehensive doctrines. It must be made explicit that a contracting identity takes priority over other elements of a citizen’s comprehensive doctrine and, indeed, the clearer we can be about the considerations that it excludes and the circumstances in which it takes priority, the better.

Second, I suggested that, as a species of ideal theory, the social contract could be defended by showing how an ideal of the person can guide action in non-ideal circumstances. Although an ideal theory will not spit out a definite policy for every eventuality, it matters that it can identify obligations and obligatory ends in a way that leads individuals to identify against impermissible strategies and unacceptable circumstances. The feasibility conditions of a useful theory of justice are sensitive to the ways in which people conceive of themselves and this is partly set by the hierarchy of principles to which they are committed and the further principles that follow from them.

There are, I think, a number of ways in which these insights are already well served by prominent elements of contract theory. In particular, publicity conditions ensure that the justification for the principles of justice is widely promulgated to citizens and the primary good of self-respect is designed to buttress citizens’ sense of self-worth so that they can see their values reflected in public institutions and proceed on the basis that their lives and plans are respected by society. The rules of public reason that Rawls (2005, pp.212-254; pp.435-490) outlines are also instructive for thinking about how a complementary identity-based approach to a contract theory might fill in some of the detail of what it means to actually have a sense of justice. In any case, I hope to have shown in this chapter that inverting the contract would make a useful addition to the contract theorist’s arsenal.

# *Conclusion*

In this dissertation I have sought to capture the binding force of obligation and to explain what it means to be under a duty. My contention has been that only on a sanction theory can we generate the necessary conditions for an individual to actually feel the sense of restraint that is characteristic of obligation. However, not just any sanction will do. It must have sufficient severity to deliver a meaningful penalty if incurred but it must also recede into the background in our deliberations so that our relationship to our reasons is not skewed by fear or self-interest.

There is a sanction that fits the bill. To violate a principle to which you are committed as part of how you think of yourself is to unpick your own identity. A principle forms a part of a person’s identity when she assigns it the power to exclude some considerations from counting as reasons for her under certain conditions. To act on an excluded consideration and violate a principle through which you conceive of yourself is to make yourself unintelligible. If your actions do not match your principles then who are you? This is a serious cost to bear since it matters profoundly to almost all of us that we are somebody in particular, somebody who can be defined by the specific hierarchy of principles to which he is committed.

However, the way this sanction works is not primarily by instilling fear. Rather, the concern to not be nobody is really only the negative form of the interest we take in being somebody. We aspire to construct identities that express the reasons that we acknowledge as we go on about the business of living our lives. One important way that we can respond to those reasons is to make rules for ourselves that form part of how we understand ourselves. Even though the presence of a sanction makes it unacceptable, in a very real sense, to breach our commitments, we can be said to impose this cost upon ourselves voluntarily if we endorse our principles because of the reasons that count in their favour, that is if they seem to us a sufficiently justified method of realising and respecting our values.

An obligation, then, can only be said to exist when a principle is morally justified *and* a part of an individual’s identity. Moral justification underpins judgments of liability and blame, but bindingness is only generated by the special relationship between a person and her principles. This position explains how it is possible to do your duty for the right reasons even though you can feel the weight of duty pressing down. Further, it permits reasonable hope that we can, at least sometimes, attain a state of grace when we attend only to the relevant moral considerations and have no need of the thought that anything but the right action would undermine our identities.

My exploration of obligation and bindingness was prompted by the failure of a broadly Rawlsian social contract model to satisfactorily establish that we are bound to obey suitably just states. I looked at two approaches to political obligation. The first, the sufficiency thesis, holds that sufficient moral justification can generate an obligation to obey the state while the second, the insufficiency thesis, denies this and claims instead that some manner of voluntary participation is required if the state is to be permitted to rule over you.

In the end, I rejected both. The insufficiency thesis is too insensitive to the demands of justice and the potential virtues of worthy states. The sufficiency

thesis, on the other hand, runs into trouble because it neglects the strong intuition that substantive voluntary participation is pivotal for the bindingness of our obligations to hold. In fact, it offered no answer to the question of what it means for an obligation to be binding at all and, as if that was not enough, it gutted the contract tradition by emasculating its traditional concern for the ideal of voluntariness.

Plugging in my identity-based sanction theory of the binding force of obligation solved these problems and facilitated something else as well. It opened up the conceptual space for hypothetical consent to play a meaningful role in a political philosophy. Establishing that a person would have agreed to some principle is not a superficial achievement. It invites that person to acknowledge the principle as binding upon her or to damage her self-conception because the only way to reject the principle is to forsake those other commitments and values that would have moved her to embrace it.

This insight formed the basis of a relational theory of legitimacy. For the sake of the efficacy and stability, and therefore the moral justification, of a just political authority it will almost certainly matter that a proportion of its citizens feel bound to obey it. It is a condition of a state’s legitimacy, then, that it has the power to tap into the binding force of obligation by extending the practical identities of citizens around new policies and laws. However, just states also play a meaning-giving role for those individuals whose practical identities support the authority of the state. Putting these ideas together, I submit that legitimacy is best understood as a binary relationship between states and individual citizens when the state is morally justified and empowered to extend a particular citizen’s practical identity.

Voluntariness is important here because identifying with principles is an effective and elegant strategy for responding to our reasons. However, it is only when we endorse a principle because of the considerations that make it acceptable to us that we express our values in our responses to reasons. States serve our interests in this regard only when we voluntarily commit to them. For this reason, and for the sake of stability and efficacy, states should work hard to make citizens’ identities and the principles that follow from them transparent.

Identification with principles is one critical element in the conception of autonomy I developed from the interest, that I take almost all of us to have, in living an intelligible life. I argued that personal autonomy and identification with principles are two expressions of that interest and that together these three conceptions of what it means to be an autonomous agent constitute a uniquely worthwhile form of human existence.

With these ideas in hand, I suggested a renovation of the traditional social contract model to include a complementary strategy. I argued that in addition to figuring out those principles to which it would be sensible to agree, we need to rationally reconstruct the prospective citizen in order to determine what a person must be like in order to relate to a just state in an obligatory way. Only then will we have a complete picture; the just state embraced by the conscientious citizen. This model was of some use in explaining how stability can be achieved and in vindicating the significance of ideal theory in the limits that it sets on the identities that non-ideal compromises can be allowed to promote.

To conclude then, social contract theorists must refocus on their conception of the person and acknowledge the essential role that identity and identity formation play in binding individuals to those principles and institutions that can be morally justified. If they do this, then it will be open to them to reclaim hypothetical consent and use it to show how we must relate, on pain of unintelligibility, towards the conclusions that they reach.

More generally, I hope to have shown that the binding force of obligation is profitably thought of as a very particular, and effective, method of responding to reasons. In fact, it is by conceiving of ourselves through our commitment to various principles that we can respond to reasons in a comprehensive manner. And that is by living an intelligible life.

***Bibliography***

Adams, R. M. (1999) *Finite and Infinite Goods: A Framework for Ethics.* Oxford: Oxford University Press.

Anscombe, G. E. M. (1958) ‘Modern Moral Philosophy’, *Philosophy*, 33 (124), pp.1-19.

Aristotle (1992) *The Politics.* trans. by Sinclair, T. A. London: Penguin.

––––– (2009) *The Nicomachean* *Ethics.* trans. by Ross, D. Oxford: Oxford University Press.

BBC (2oo5) *North Korea wages war on long hair*. Available at: <http://news.bbc.co.uk/1/hi/world/asia-pacific/4157121.stm> (Accessed: 26/3/14).

––––– (2014) *North Korea: Students required to get Kim Jong-un haircut*. Available at: <http://www.bbc.co.uk/news/blogs-news-from-elsewhere-26747649> (Accessed: 26/3/14).

Beran, H. (1987) *The Consent Theory of Political Obligation*. London: Croom Helm.

Baier, K. (1989) ‘Justice and the Aims of Political Philosophy’, *Ethics*,99 (4), pp.771-790.

Baron, M. W. (1995) *Kantian Ethics Almost Without Apology*. Ithaca, NY: Cornell University Press.

Barry, B. (1995) ‘John Rawls and the Search for Stability’, *Ethics*, 104 (5), pp.874-915.

Berlin, I. (1969) ‘Two Concepts of Liberty’, in Berlin, I. Four Essays on Liberty. Oxford: Oxford University Press.

––––– (1981) *Against the Current: Essays in the History of Ideas*. Oxford: Clarendon Press.

Brandt, R. (1964) ‘The Concepts of Obligation and Duty’, *Mind*, 73, pp.374-93.

Bratman, M. E. (2001) ‘Two Problems about Human Agency’, *Proceedings of the Aristotelian Society*, 101, pp.309-326.

––––– (2007) *Structures of Agency: Essays*. Oxford: Oxford University Press.

Brettschneider, C. (2012) *When the State Speaks, What Should it Say? How Democracies Can Protect Expression and Promote Equality* Princeton, NJ: Princeton University Press.

Buchanan, A. (2002) ‘Political Legitimacy and Democracy’, *Ethics*, 112, pp.689-719.

Butler, J. (1970) *Fifteen Sermons Preached at the Rolls Chapel and a Dissertation of the Nature of Virtue*. Roberts, T. A. (ed.) London: Society for the Preservation of Christian Knowledge.

Callan, E. (1997) *Creating Citizens: Political Education and Liberal Democracy*. Oxford: Oxford University Press

Caney, S. (2014) ‘Two Kinds of Climate Justice: Avoiding Harms and Sharing Burdens’, *The Journal of Political Philosophy*, 22 (2), pp.125-149.

Christiano, T. (1996) The Rule of the Many: Fundamental Issues in Democratic Theory. Boulder, CO: Westview Press.

––––– (2004) ‘The Authority of Democracy’, Journal of Political Philosophy, 12 (3), pp.266–290.

Christman, J. (1991a) ‘Autonomy and Personal History’, *Canadian Journal of Philosophy*, 21 (1), pp.1-24.

––––– (1991b) ‘Liberalism and Individual Positive Freedom’, *Ethics*, 101, pp.343-359.

Cohen, G. A. (1996) ‘Reason, Humanity, and the Moral Law’, in O’Neill, O. (ed.) *The Sources of Normativity*.Cambridge: Cambridge University Press, pp.167-188.

––––– (2005) ‘Facts and Principles’, *Philosophy & Public Affairs*, 31 (3), pp.211-245.

––––– (2008) *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.

Cohen, J. (1989) ‘Deliberation and Democratic Legitimacy’, in Hamlin, A. and Pettit, P. (eds.) *The Good Polity: Normative Analysis of the* *State*. New York, NY: Blackwell.

Colburn, B. (2008) ‘The Concept of Voluntariness’, *The Journal of Political Philosophy*, 16 (1), pp.101-111.

––––– (2010) *Autonomy and Liberalism*. New York, NY: Routledge.

Darwall, S. L. (1977) ‘Two Kinds of Respect’, *Ethics*, 88, pp.36-49.

––––– (1983) *Impartial Reason*. New York, NY: Cornell University Press.

––––– (2004) ‘Respect and the Second-Person Standpoint’, *Proceedings of the American Philosophical Association*, 78, pp.43-60.

––––– (2010) ‘Authority and Reasons: Exclusionary and Second Personal’, *Ethics*, 120 (2), pp.257-278.

––––– (2009) *The Second Person Standpoint: Morality, Respect, and Accountability.* Cambridge, MA: Harvard University Press.

Dworkin, G. (1988) *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Dworkin, R. (1975) ‘The Original Position’, in Daniels, N. (ed.) *Reading Rawls*. New York, NY: Basic Books.

Edmundson, W. A. (1998) ‘Legitimate Authority without Political Obligation’, *Law and Philosophy*, 17, pp.43-60.

Farrelly, C. (2007) ‘Justice in Ideal Theory: A Refutation’, *Political Studies*, 55 (4), pp.844-864.

Feinberg, J. (1980) *Rights, Justice, and the Bounds of Liberty: Essays in Social Philosophy*. Princeton, NJ: Princeton University Press.

Foot, P. (1972) ‘Morality as a System of Hypothetical Imperatives’, *The Philosophical Review*, 81 (3), pp.305-316.

Freeman, S. (1990) ‘Reason & Agreement in Social Contract Views’, *Philosophy & Public Affairs*,19 (2), Spring, pp.122-157.

Frankfurt, H. (1971) ‘Freedom of the Will and the Concept of a Person’, *The Journal of Philosophy*, 68 (1), pp.5-20.

––––– (1988) *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.

Friedman, M. (2000) ‘Autonomy, Social Disruption, and Women’, in Mackenzie, C. and Stoljar, N. (eds.) *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford: Oxford University Press, pp. 35-51.

Gallie, W.B. (1955) ‘Essentially Contested Concepts’, *Proceedings of the Aristotelian Society*,56, pp.167-198.

Gardner, J. (2007) *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*. Oxford: Oxford University Press.

Gaus, G. (2013) ‘Why the Conventionalist Needs the Social Contract (and Vice Versa)’, *Rationality, Markets, Morals*, 4, pp.71-87.

Gauthier, D. (1986) *Morals By Agreement*. Oxford: Clarendon Press.

Green, L. (1988) *The Authority of the State*. Oxford: Clarendon Press.

Gutmann, A. (1999) *Democratic Education*. Princeton, NJ: Princeton University Press.

Habermas, J. (1979) *Communication and the Evolution of Society*. London: Heinemann.

––––– (1984*) Legitimation Crisis*. London: Heineman.

––––– (1997) ‘Popular Sovereignty as Procedure’, inBohman, J. and Rehg, W. (eds.) *Deliberative Democracy*. Cambridge: Cambridge University Press.

Hacker, P. M. S. (1973) ‘Sanction Theories of Duty’, inSimpson, A. W. B. (ed.) *Oxford Essays in Jurisprudence*. Oxford: Clarendon Press, pp.131-170.

Hampton, J. (1980) ‘Contracts and Choices: Does Rawls Have a Social Contract Theory?’, *Journal of Philosophy*, LXXVII (6), pp.315-338.

––––– (1986) *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.

Hart, H. L. A. (1994) *The Concept of Law: Second Edition*. Oxford: Oxford University Press.

Hegel, G.W. F. (2010) *Elements of the Philosophy of Right*. Cambridge: Cambridge University Press.

Hieronymi, P. (2005) ‘The Wrong Kind of Reason’, *The Journal of Philosophy*, 102 (9), pp.437-457.

Hobbes, T. (1985) *Leviathan*. London: Penguin.

Hurd, H. M. (1996) ‘The Moral Magic of Consent’, *Legal Theory*, 2 (2), pp.121-146.

Hurka, T. (2011) *Drawing Morals: Essays in Ethical Theory*. Oxford: Oxford University Press.

Hurley, S. L. (1989) *Natural Reasons: Personality and Polity*. Oxford: Oxford University Press.

Joyce, R. (2001) *The Myth of Morality*. Cambridge: Cambridge University Press.

Kagan, S. (1989) *The Limits of Morality*. Oxford: Clarendon Press.

Kant, I. (1996) *The Metaphysics of Morals*. trans. by Gregor, M. Cambridge: Cambridge University Press.

––––– (1998) *The Groundwork of the Metaphysics of Morals*. trans. by Gregor, M. Cambridge: Cambridge University Press.

Kennett, J. (1993) ‘Mixed Motives’, *Australasian Journal of Philosophy*, 71 (3), pp.256-269.

Klosko, G. (1990) ‘The Obligation to Contribute to Discretionary Public Goods’, *Political Studies*, XXXVIII, pp.196-214.

––––– (2005) *Political Obligations*. Oxford: Oxford University Press.

Knight, J. and Johnson, J. (1994) ‘Aggregation and Deliberation: On the Possibility of Democratic Legitimacy’, *Political Theory*, 22 (2), pp.277-296.

Knowles, D. (2002) *Hegel and the Philosophy of Right*. London: Routledge.

––––– (2010) *Political Obligation: A Critical Introduction*. Abingdon: Routledge.

Korsgaard, C. M. (1996a) ‘The Sources of Normativity’, in O’Neill, O. (ed.) *The Sources of Normativity*. Cambridge: Cambridge University Press.

––––– (1996b) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Ladenson, R. (1980) ‘In Defense of a Hobbesian Conception of Law’, Philosophy & Public Affairs, 9 (2), pp.134-159.

Lenman, J. (2009) ‘The Politics of the Self: Stability, Normativity and the Lives we can Live with Living’, in Bortolotti, L. (ed.) *Philosophy and Happiness*. London: Palgrave Macmillan, pp.183-199.

Locke, J. (2004) *Two Treatises of Government*. London: Everyman.

MacIntyre, A. (2007) *After Virtue*. London: Duckworth.

Mackie, J. L. (1990) *Ethics: Inventing Right and Wrong*. London: Penguin.

Martin, R. (1975) ‘Two Models for Justifying Political Authority’, *Ethics*, 86 (1), pp.70-76.

McDowell, J. (1996) *Mind and World*. Cambridge, MA: Harvard University Press.

MacCallum, G. C. (1967) ‘Negative and Positive Freedom’, *Philosophical Review*, 76, pp.312-334.

McTernan, E. (2014) ‘How to Make Citizens Behave: Social Psychology, Liberal Virtues, and Social Norms’, *The Journal of Political Philosophy*,22 (1), pp.84-104.

Mill, J. S. (1979) *Utilitarianism*. Glasgow: Collins.

Miller, D. (2011) ‘Political Philosophy for Earthlings’, in Leopold, D. and Stears, M. *Political Theory: Methods and Approaches*. Oxford: Oxford University Press, pp.29-48.

Nagel, T. (1959) ‘Hobbes's Concept of Obligation’, Philosophical Review, 68, pp.68-83.

––––– (1986) *The View From Nowhere*. Oxford: Oxford University Press.

––––– (1987) ‘Moral Conflict and Political Legitimacy’, *Philosophy & Public Affairs*, 16 (3), Summer, pp.215-240.

––––– (1991) *Equality and Partiality*. Oxford: Oxford University Press.

––––– (1996) ‘Universality and the Reflective Self’, in O’Neill, O. (ed.) *The Sources of Normativity*.Cambridge: Cambridge University Press, pp.200-209.

Neuhouser, F. (2000) *Foundations of Hegel’s Social Theory: Actualizing Freedom*. Cambridge, MA: Harvard University Press.

Nietzsche, F. (1967) *The Will to Power.* trans. by Kaufmann, W. and Hollingdale, R. J. New York, NY: Vintage Books.

Nozick, R. (1974) Anarchy, State, and Utopia. New York, NY: Basic Books.

Nussbaum, M. (2000) *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

Olsaretti, S. (2004) *Liberty, Desert, and the Market*. Cambridge: Cambridge University Press.

––––– (2008) ‘The Concept of Voluntariness – A Reply’, *The Journal of Political Philosophy*, 16 (1), pp.112-121.

––––– (2009) ‘Responsibility and the Consequences of Choice’, *Proceedings of the Aristotelian Society*, Vol. CIX, (2), pp.165-188.

O’Neill, O. (2003a) ‘Autonomy: The Emperor’s New Clothes’, Aristotelian Society Supplementary Volume, 77 (1), pp.1–20.

––––– (2003b) ‘Constructivism vs. Contractualism’, *Ratio (new series)* XVI, (4), pp.319-331.

Oshana, M. (1998) ‘Personal Autonomy and Society’, Journal of Social Philosophy, 29 (1), pp.81–102.

Parfit, D. (1987) *Reasons and Persons*. Oxford: Clarendon Press.

Pettit, P. (1996) ‘Freedom as Antipower’, *Ethics*, 106, pp.576-604.

Pitkin, H. (1965) ‘Obligation and Consent – I.’, *The American Political Science Review*, 59, December, pp.990-999.

––––– (1966) ‘Obligation and Consent – II.’, *The American Political Science Review*, 60, March, pp.39-52.

Plato (2005) ‘Republic’, trans. by Shorey, P. in Hamilton, E. and Cairns, H. (eds.) *Plato: The Collected Dialogues*. Princeton, NJ: Princeton University Press.

Prichard, H. A. (1912) ‘Does Moral Philosophy Rest on a Mistake?’, *Mind*, 21, pp.21-37.

Rawls, J. (1964) ‘Legal Obligation and the Duty of Fair Play’, in Hook, S. (ed.) *Law and Philosophy: A Symposium*. New York, NY: New York University Press.

––––– (1980) ‘Kantian Constructivism in Moral Theory’, *Journal of Philosophy*, 77 (9), pp.515-572.

––––– (1985) ‘Justice as Fairness: Political not Metaphysical’, *Philosophy and Public Affairs*,14 (3), pp.223-351.

––––– (1999) *A Theory of Justice: Revised Edition*. Cambridge, MA: Harvard University Press.

––––– (2000) *Lectures on the History of Moral Philosophy*. Cambridge, MA: Harvard University Press.

––––– (2003) *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

––––– (2005) *Political Liberalism: Expanded Edition*. New York, NY: Columbia University Press.

Raz, J. (1985) ‘Authority and Justification’, *Philosophy and Public Affairs*, 14 (1), pp.3-29.

––––– (1986) *The Morality of Freedom*. Oxford: Clarendon Press.

––––– (1990) *Practical Reasons and Norms*. Princeton, NJ: Princeton University Press.

––––– (2006) ‘The Problem of Authority: Revisiting the Service Conception’, *Minnesota Law Review*, 90, pp.1003-1044.

––––– (2010) ‘On Respect, Authority, and Neutrality: A Response’, *Ethics*, 120 (2), pp.279-301.

Ross, W. D. (1930) *The Right and the Good*. Oxford: Oxford University Press.

Rousseau, J.J. (1984) *A Discourse on Inequality*. London: Penguin.

––––– (1986) *The Social Contract*.London: Penguin.

Sandel, M. (1984) ‘The Procedural Republic and the Unencumbered Self’, *Political Theory*, 12 (1), pp.81-96.

Sen, A. (2009) *The Idea of Justice*. Cambridge, MA: Harvard University Press.

Scanlon, T. M. (1972) ‘A Theory of Freedom of Expression’, Philosophy & Public Affairs, 1 (2), pp.204-226.

––––– (1976) ‘Nozick on Rights, Liberty, and Property’, *Philosophy & Public Affairs*, 6 (1), Autumn, pp.3-25.

––––– (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schier, F. (1993) ‘The Kantian Gulag: Autonomy and the Liberal Conception of Freedom’, in Knowles, D. and Skorupski, J. (eds.)*Virtue and Taste: Essays on Politics, Ethics & Aesthetics*. London: Blackwell.

Schiller, F. (2005) ‘On Grace and Dignity’, in Curran, J. V. and Fricker, C. (eds.) *Schiller’s ‘On Grace and Dignity’ in its Cultural Context: Essays and a New Translation*. New York, NY: Camden House.

Schneewind, J. B. (1998) *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.

Shaw, T. (2010) *Nietzsche’s Political Skepticism*. Princeton, NJ: Princeton University Press.

Simmons, A. J. (1979) *Moral Principles and Political Obligations*. Princeton, NJ: Princeton University Press.

––––– (1987) ‘The Anarchist Position: A Reply to Klosko and Senor’, *Philosophy & Public Affairs*, 16 (3), Summer, pp.269-279.

––––– (1993) ‘Liberal Impartiality and Political Legitimacy’, *Philosophical Books*, 34 (4), pp.213-223.

––––– (1999) ‘Justification & Legitimacy’, *Ethics,* 109, pp.739–771.

––––– (2010) ‘Ideal and Nonideal Theory’, *Philosophy & Public Affairs*, 38 (1), pp.5-36.

Stark, C. (2000) ‘Hypothetical Consent & Justification’, *The Journal of Philosophy*, 97 (6), pp.313–334.

Stern, R. (2012) *Understanding Moral Obligation: Kant, Hegel, Kierkegaard*. Cambridge: Cambridge University Press.

––––– (2013) ‘Kant, Moral Obligation, and the Holy Will’, inTimmons, M. and Baiasu, S. (eds.) *Kant on Practical Justification: Interpretive Essays*. Oxford: Oxford University Press, pp.125-152.

Tadros, V. (2011) ‘Consent to Harm’, *Current Legal Problems*, 64, pp.23-49.

Taylor, C. (1979) ‘What’s Wrong With Negative Liberty?’, in Ryan, A. (ed.) *The Idea of Freedom*. Oxford: Oxford University Press.

––––– (1989) *Sources of the Self: The Making of the Modern Identity*.Cambridge: Cambridge University Press.

––––– (1991) *The Ethics of Authenticity*. Cambridge, MA: Harvard University Press.

Tiberius, V. (2000) *Deliberation about the Good : Justifying What We Value*. New York, NY: Garland.

––––– (2008) *The Reflective Life : Living Wisely with Our Limits*. Oxford: Oxford University Press.

Timmons, M. (2003) ‘The Limits of Moral Constructivism’, *Ratio (new series),* XIV, (4), pp.391-423.

Tsai, G. (2014) ‘Rational Persuasion as Paternalism’, *Philosophy & Public Affairs*, 42 (1), pp. 78-112.

Valentini, L. (2012) ‘Ideal vs. Non-ideal Theory: A Conceptual Map’, *Philosophy Compass*, 7 (9), pp.654-664.

Von Hayek, F. A. (1991) ‘Freedom and Coercion’, in Miller, D. (ed.) *Liberty*. Oxford: Oxford University Press

Waldron, J. (1987) ‘Theoretical Foundations of Liberalism’, *Philosophical Quarterly*,37 (147), April, pp.127-150.

––––– (1993) ‘Special Ties and Natural Duties’, *Philosophy & Public Affairs*, 22 (1), Winter, pp.3-30.

Waskan, Jonathan (1998) ‘De Facto Legitimacy and Popular Will’, *Social Theory and Practice*, 24 (1), pp.25-56.

Warrender, H. (1957) The Political Philosophy of Hobbes: his Theory of Obligation. Oxford: Oxford University Press.

Weber, M. (1964) *The Theory of Social and Economic Organization.* London: Collier McMillan.

––––– (2004) *The Vocation Lectures*. trans. by Livingstone, R. Indianapolis IN: Hackett.

Wellman, C. H. (1996) ‘Liberalism, Samaritanism, and Political Legitimacy’, *Philosophy & Public Affairs*, 25, pp.211-237.

––––– (2005) ‘Samaritanism and the Duty to Obey the Law’, inWellman, C. H. and Simmons, A. J. (eds.) *Is There a Duty to Obey the Law?* Cambridge: Cambridge University Press.

Wilde, O. (1988) *The Complete Plays*. London: Methuen.

Williams, B. (1999) *Moral Luck*. Cambridge: Cambridge University Press.

Wolf, S. (2009) ‘Moral Obligations and Social Commands’, in Newlands, S. and Jorgensen, L. M. (eds.) *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*. Oxford: Oxford University Press, pp.343-367.

Wolff, R. P. (1998) *In Defense of Anarchism*. Berkeley, CA: University of California Press.

1. This is Dworkin’s (1975) ‘standard indictment’ of hypothetical consent. I will discuss it in Section 1.4. [↑](#footnote-ref-1)
2. For a discussion of the relationship between natural law and obligation in Hobbes, and whether he is offering a theory of moral obligation, see Nagel (1959) and Warrender (1957). Regarding Hobbes himself(e), see (1985, esp. Ch.XVIII). [↑](#footnote-ref-2)
3. Although it seems correct to say that ‘obligation’ originally implied some variety of actual consent while ‘duty’ was used to convey a non-optional relationship of obedience or requirement, I intend to follow Brandt (1964) and Knowles (2010) in taking the concepts of duty and obligation to mean generally the same thing in their contemporary usage. Rawls, of course, distinguishes between them precisely in these terms (1999, p.98; p.296) but in this he is very much the exception rather than the rule. [↑](#footnote-ref-3)
4. This is not to say that my suggestions here may not have wider relevance, indeed I hope they do. [↑](#footnote-ref-4)
5. See Freeman (1990, pp.140-156) for a discussion of the role of agreement in a Rawlsian scheme. Freeman (1990, p.135) does suggest that we should understand the contract as a heuristic. However, insofar as he follows Rawls’s deeper constructivism about the legitimacy of moral principles it would appear that he too gives the ideal of agreement a more substantial role. [↑](#footnote-ref-5)
6. For the veil of ignorance see Rawls (1999, pp.10-24). For further discussion of the idealisation of the representative parties see Section 6.2. [↑](#footnote-ref-6)
7. Actual consent theories are hamstrung by the practical impossibility of arranging for everyone (or nearly everyone) to sign up the agreement. See Beran (1987) for a discussion of the limitations of actual consent as a theory of political obligation. [↑](#footnote-ref-7)
8. See Section 1.4. [↑](#footnote-ref-8)
9. See Gaus (2013) for a conventionalist use of the social contract. [↑](#footnote-ref-9)
10. See Waskan (1998) for a discussion of the relevance of broader issues that are often included in debates about ‘legitimacy’. [↑](#footnote-ref-10)
11. Nozick (1974, p.134) introduces a similar distinction when discussing entitlement: “we must distinguish between an agency’s being entitled to be the one wielding certain power from its being entitled to wield that power”. [↑](#footnote-ref-11)
12. For a discussion of the vagaries and problems of tacit consent see Simmons (1979, pp.75-100). [↑](#footnote-ref-12)
13. As in the principle of fair play. See Rawls (1964). [↑](#footnote-ref-13)
14. See Waskan (1998) for some criteria that a state must meet to demonstrate its legitimacy. [↑](#footnote-ref-14)
15. Indeed, Buchanan (2002, p.693) concludes that “neither Rawls nor Raz find the question of political authority (as including the right to be obeyed) to be of much consequence for political theory”. I will discuss Buchanan’s ideas about legitimacy in detail in Chapter 4. [↑](#footnote-ref-15)
16. I will suggest a complementary goal for contract theorists in Chapter 6. [↑](#footnote-ref-16)
17. See Simmons (1979, p.31-35) for a more thorough exposition of the particularity requirement. [↑](#footnote-ref-17)
18. In Chapters 2 and 3 I will argue that how a person thinks of herself and how her actions match up to her identity are crucial to her capacity to feel the binding force of obligation. For a discussion of the “moral magic of consent,” see Hurd (1996). [↑](#footnote-ref-18)
19. In the long-term at any rate. Despots are often welcomed for the short-term solutions they offer. [↑](#footnote-ref-19)
20. We might well still feel bound by promises made to illegitimate authorities. This is why I shall argue in Chapter 4 that only a morally justified state can be a legitimate one. [↑](#footnote-ref-20)
21. “The temptation is to say that an institution ‘applies’ to me only if I have voluntarily brought myself under its auspices, or to impose some other similar condition (such as receipt of benefits) on any inference from the justice of the institution to a duty of obedience” (Waldron 1993, p.7). [↑](#footnote-ref-21)
22. Waldron adopts a Kantian story about how the state of nature compels people to give up their natural freedoms in order to secure their claims to property and resources and to provide a guarantee against violence and instability. See Kant (1996, pp.89-113/6:311-6:342). [↑](#footnote-ref-22)
23. Stark intends to cover contract theories based on self-interest as well as the ‘contractualist’ theories that introduce some moral content from the start. For a discussion of the applicability and utility of the ‘contractarian’ and ‘contractualist’ labels see O’Neill (2003b). For some questions about Rawls’s social contract credentials see Hampton (1980). For a helpful discussion of how contract theory intersects with ‘constructivism’ in all of its many forms, i.e. political, moral, and metaethical, see Timmons (2003). [↑](#footnote-ref-23)
24. See Rawls (1999, pp.330-331). [↑](#footnote-ref-24)
25. For a discussion of the veracity of Nagel’s commitment to ‘voluntariness’ see Simmons (1993, p.220-222). [↑](#footnote-ref-25)
26. Either express or tacit. I side with Simmons (1979) in his categorisation of Locke as an actual rather than hypothetical consent theorist, although nothing hangs on that here. [↑](#footnote-ref-26)
27. “Men being, as has been said, by Nature, all free, equal and independent, no one can be put out of this Estate, and subjected to the Political Power of another, without his own *Consent*. The only way whereby any one divests himself of his Natural Liberty, and puts on the bonds of Civil Society is by agreeing with other Men to joyn and unite into a Community, for their comfortable, safe, and peaceable living one amongst another, in a secure Enjoyment of their Properties, and a greater Security against any that are not of it” (Locke 2004, p.331). See also Locke (2004, p.332): “And thus every Man, by consenting with others to make one Body Politick under one Government, puts himself under an Obligation to every one of that Society, to submit to the determination of the *majority*, and to be concluded by it; or else this *original Compact*, whereby he with others incorporates into *one Society*, would signifie nothing, and be no Compact, if he be left free, and under no other ties, than he was in before in the State of Nature”. [↑](#footnote-ref-27)
28. Indeed, a philosophical anarchist might hold that this is the closest we can come to showing that we could have obligations to obey states. [↑](#footnote-ref-28)
29. For such a theory see Wolff (1998, esp. pp.18-21). [↑](#footnote-ref-29)
30. For a helpful discussion of the justification of coercion by the demands of justice see Buchanan (2002, esp. pp.703-709). [↑](#footnote-ref-30)
31. Knowles (2010, p.25) prefers to characterise such rights as liberty-rights to get the point across that the right-bearer has no duty to refrain from doing something. I prefer “justification-right” as a term because I think that it better encapsulates the key notion that reasons sufficient to justify an action can give it a positive moral standing. In the Hohfeldian terminology that Knowles prefers, we would have to add something like a second-order immunity from interference or subsequent prosecution to fill out the idea. Even so, the meaning would not be quite the same so I will adopt Ladenson’s lexicon. [↑](#footnote-ref-31)
32. See Section 1.3. [↑](#footnote-ref-32)
33. See Sections 4.4 and 4.5 for an extended discussion of this claim. I argue there that since efficacy is an important condition of the justification of a state it matters that it can deploy the binding force of obligation. [↑](#footnote-ref-33)
34. I will consider Raz’s account of authority in the next chapter. See also Section 5.5 where I argue that identifying with principles, such as the principle to obey just states, is an important expression of autonomy. [↑](#footnote-ref-34)
35. We can link this account to the earlier discussion of the legitimacy of foreign states. In these terms, legitimacy might be thought to entail an obligation not to interfere or undermine the workings of a legitimate state to which we are not subject. [↑](#footnote-ref-35)
36. Of course, there are conditions of legitimacy that define the scope of any agent’s authority. If they act *ultra vires* then we are entitled to question their orders. I will discuss this in more detail when I come to Raz’s theory of exclusionary reasons in Chapter 2. [↑](#footnote-ref-36)
37. See Weber (1964) for his three sources of descriptive legitimacy: law, tradition, and charisma. See also Buchanan (2002) for an argument in favour of the priority of the normative conception. [↑](#footnote-ref-37)
38. This is the natural way to interpret Rawls in particular. Recall that he (1999, p.455) says: “There is no violation of our autonomy so long as its principles are properly followed”. Similarly, Nagel (1991, p.36) asserts that: “To show that they all have sufficient reason to accept it is as close as we can come to making this involuntary condition voluntary”. [↑](#footnote-ref-38)
39. See Rawls (2005, pp.77-81). [↑](#footnote-ref-39)
40. Stark (2000, pp.318-319) does acknowledge the general problem of bindingness, but does not attempt to delve any deeper. My contention here is that this is a critical oversight. [↑](#footnote-ref-40)
41. I will discuss the relationship between the representative parties and real people in more detail in Chapter 6. [↑](#footnote-ref-41)
42. “Oh, duty is what one expects from others, it is not what one does oneself” (Wilde 1988, p.339). The character that says it is Lord Illingworth from *A Woman of No Importance*. [↑](#footnote-ref-42)
43. See Brandt (1964) for a discussion of the interchangeability of duty and obligation. Hacker (1973) traces this trend as far back as Bentham. In this chapter I will, for ease of exposition and the sake of variety, use them as synonyms. [↑](#footnote-ref-43)
44. I shall assume a broadly constructivist story about the justification of specific moral principles but I hope that my account of the binding force of obligation will have wider application. [↑](#footnote-ref-44)
45. See, for example, Simmons (1979), Beran (1987), Klosko (2005), and Knowles (2010). [↑](#footnote-ref-45)
46. See Dworkin (1975, p.17-21). [↑](#footnote-ref-46)
47. Simmons’s interpretation of Locke is quite sophisticated. He understands him as an actual consent theorist, but suggests that Locke may also have strayed into other grounds for political obligation in his discussion of tacit consent. See Simmons (1979) for a detailed discussion. [↑](#footnote-ref-47)
48. This is not to say that principles established via hypothetical consent cannot refer to actual consent. There may well be certain conditions of legitimacy on political obligations that will only come due if those conditions are met. For instance, a morally justified state is likely to be one that treats its citizens well and one part of this might involve seeking the actual consent of the people wherever possible. I will discuss this subject in more detail in Chapter 4. [↑](#footnote-ref-48)
49. For the classic discussion of internal and external reasons see Williams (1999). [↑](#footnote-ref-49)
50. See Chapter 6 where I discuss Rawls’s account of stability and apply my proposal to explain how an ideal theory can generate the motivational resources to privilege reasons of justice in our everyday deliberations. [↑](#footnote-ref-50)
51. Recall that Nagel (1991, p.36) explains the search for legitimacy as “an attempt to realize some of the values of voluntary participation, in a system of institutions that is unavoidably compulsory”. See Chapter 4 for my account of political legitimacy. [↑](#footnote-ref-51)
52. The content of some of our more particular moral obligations can only be fixed by the actual commitments that we make, in cases such as promising, or the actual circumstances in which other people are entitled to make demands of us. However, on the broadly constructivist picture that I am presupposing here, there are more general principles to which we would agree such as, for example, the principle that we should keep our promises. [↑](#footnote-ref-52)
53. See the discussion of Ladenson’s (1980) position in Section 1.5. He points to such examples as self-defence, defence of others, and parental authority. [↑](#footnote-ref-53)
54. Arguably, Nietzsche thought that proper-functioning oughts are all there is to morality, although he is sometimes scornful of the helpfulness of such a paradigm: “A man as he ought to be: that sounds to us as insipid as ‘a tree as he ought to be’” (1967, p.181, original emphasis). For a useful discussion of Nietzsche’s approach to political and moral obligation see Shaw (2010). [↑](#footnote-ref-54)
55. “It is a desideratum of an account of the concept of duty that the varieties of duty be generically related” (Hacker 1973, p.141). Hacker does go on to urge caution on this point since there may be weightier theoretical considerations that ultimately count against this desideratum. [↑](#footnote-ref-55)
56. Kagan (1989, p.17, fn.1) makes a strong case that Ross would be better off contrasting *pro tanto* duties with all-things-considered duties. [↑](#footnote-ref-56)
57. For an interesting argument in favour of self-regarding duties see Tadros (2011). [↑](#footnote-ref-57)
58. An issue that intersects with this one is the distinction that Scanlon draws (and Olsaretti endorses) between attributing an action to a person so that we can praise or blame her and assigning substantive responsibility so that she is liable to bear any costs that fall out of her action. See Scanlon (1998, Ch.6) and also Olsaretti (2008, pp.116-119). [↑](#footnote-ref-58)
59. See Section 3.6 where I discuss grace in detail. See also Schiller (2005). [↑](#footnote-ref-59)
60. Scanlon (1998, pp.29-41) discusses desire using the term “attention-directedness”. Although my focus here is on obligation, my approach has been influenced by the emphasis he puts on how individuals relate to the options that are open to them. [↑](#footnote-ref-60)
61. For an emphasis on this question see Mill (1979, Ch.5). [↑](#footnote-ref-61)
62. The idea that morality has to do with how we understand ourselves and our relationship to reasons is not by any stretch of the imagination an original one. See Darwall (2009), Hurley (1989), and Taylor (1989; 1991) for just three thinkers who argue in a similar vein. [↑](#footnote-ref-62)
63. There are a number of influential philosophers who directly and indirectly attack the coherence of the concept of moral obligation, Mackie (1990) and Joyce (2001) are two good examples, so why concentrate on Anscombe’s critique? The answer is that Anscombe shares my interest in the question of what makes obligations stick in the distinctive way that they do. How are we to explain their binding force? [↑](#footnote-ref-63)
64. In spite of this important difference, they share the thought that our key moral terms can only be understood within the appropriate conceptual framework and that we have left that framework behind. In what follows I will focus on Anscombe’s version of the challenge. [↑](#footnote-ref-64)
65. Of course, not everybody is prepared to take God out of the picture. Anscombe is sometimes read as offering a *reductio* of that move and divine command theory is still very much alive. See Stern (2012, pp.43-48) for a good exposition of divine command theory and Adams (1999) for an interesting exponent. [↑](#footnote-ref-65)
66. There is, of course, always the nuclear option of ditching the concept and doing normative ethics without obligation. We might, as Anscombe (1958) speculates, be better off as virtue ethicists. [↑](#footnote-ref-66)
67. See Wolf (2009) for a useful taxonomy. [↑](#footnote-ref-67)
68. For the sake of the example we shall suppose that the soldier has good reasons for joining the army, there are good reasons to have armies, and good reasons for the soldier to be able to place himself under the authority of such an institution. See Raz (1990, pp.41-47) for the original example in which a soldier is ordered to appropriate a civilian’s van. [↑](#footnote-ref-68)
69. See Raz (1986, Ch.3) for his explications of the Dependency Thesis and the Normal Justification Thesis. [↑](#footnote-ref-69)
70. See Raz (1986, p.61) for his distinction between clear and unclear cases of authorities undermining their authority in this fashion and what that means for their exclusionary powers. [↑](#footnote-ref-70)
71. And why should we think that it matters more in moral cases than others? See Wolf (2009, pp.349-352) for a discussion of this question. [↑](#footnote-ref-71)
72. It is certainly true that Butler was a determined optimist. His first sermon reads at times like a Panglossian treatise on how well-suited we are to the pursuit of our own and of others’ good: “as there is no such thing as self-hatred, so neither is there any such thing as ill-will in one man towards another” (Butler 1970, p.24). He (1970, p.24) does accept that there are people “who are in great measure without the natural affections towards their fellow creatures”, but the nature of man should be judged “by what appears in the common world, in the bulk of mankind”. Even regarding our natural inclinations, our drives and appetites, his position is that when we really think them through we shall see that they can ultimately lead only to our own happiness and, indeed, the happiness of all. [↑](#footnote-ref-72)
73. If an obligation to obey the traffic laws holds under these circumstances then the driver is neither excused nor justified in ignoring the light. See Gardner (2007, esp. Ch.5) for the difference between justifications and excuses. [↑](#footnote-ref-73)
74. One might object that what is at stake in this example is respect. I fail to respect my fellow citizens by making an exception of myself with regard to a general rule in our community and for that disrespect I should feel guilty. However, I find it difficult to imagine anybody feeling disrespected in such a case. [↑](#footnote-ref-74)
75. In Chapter 5 I will argue that autonomy consists in responding to reasons and suggest that in some important cases that involves committing to a political community. [↑](#footnote-ref-75)
76. I will spend some time explaining this idea in the next section. [↑](#footnote-ref-76)
77. It sometimes seems that Butler has something not altogether dissimilar in mind when he considers the force of conscience. He describes it in terms of the approximation of an ideal. Part of that process involves appraising ourselves in comparison to God’s perfection. Butler understands the love of God to take the form of imagining a being with our capacities (that we cannot help but value) perfected, and then responding to such incomparable worth. “Religion does not demand new affections, but only claims the direction of those you already have, those affections you daily feel; though unhappily confined to objects, not altogether unsuitable, but altogether unequal to them” (Butler 1970, p.128). [↑](#footnote-ref-77)
78. See Lenman’s (2009, pp.183-185) ‘Super-Anchorites’ for an idea of what such creatures might be like. [↑](#footnote-ref-78)
79. See Scanlon (1998). [↑](#footnote-ref-79)
80. Rather than, say, evolutionary processes. [↑](#footnote-ref-80)
81. For a painstaking treatment of personal identity see Parfit (1987, esp. pp.199-244). [↑](#footnote-ref-81)
82. The ‘self’ is a vexed concept in philosophy and I will not attempt to analyse or define it here. Rather, I will just pick out some important features to help explain the overall view I am defending. For a detailed treatment of the evolution of the self and modern identity see Taylor (1989). [↑](#footnote-ref-82)
83. See Section 1.4. [↑](#footnote-ref-83)
84. The term “practical identity” is Korsgaard’s (1996a). I shall consider her account in Chapter 3. [↑](#footnote-ref-84)
85. Of course, “an account of voluntariness must avoid both rendering claims of force and of voluntariness completely subjective” (Olsaretti 2004, p.153). Olsaretti holds that unacceptability can be determined by using an objective standard of well-being, but that is for ordinary choices rather than the issues to be decided by hypothetical consent. [↑](#footnote-ref-85)
86. For an illuminating discussion of this issue see the debate between Colburn (2008) and Olsaretti (2008) on this issue. Although I do not have the luxury of exploring this issue further, one interesting possibility that merits further consideration is that a contract approach is particularly well-suited to informing people of their acceptable options when it comes to moral and political principles precisely because, as I conceive it, it extends their practical identities by deploying their hypothetical consent. [↑](#footnote-ref-86)
87. “Factually wrong beliefs about what options are available, but not wrong evaluative judgments about the nature of those options, can make someone’s choice non-voluntary” (Olsaretti 2008, p.113). [↑](#footnote-ref-87)
88. In later work, Colburn (2010, p.32) employs Olsaretti’s account of voluntariness without revision so it is not clear if he still thinks that this objection holds. [↑](#footnote-ref-88)
89. See Section 3.6 where I discuss the possibility of grace. [↑](#footnote-ref-89)
90. See Schiller (2005, p.152) on this point: “For this reason the actions of a beautiful soul are not themselves ethical, but the character as a whole is so”. [↑](#footnote-ref-90)
91. It will be important to see in Chapter 5 why voluntariness matters for autonomy. If you are conditioned to identify with some principle then you cannot impose it *because* of the reasons that count in its favour and make it a principle worth enshrining. [↑](#footnote-ref-91)
92. I shall leave aside questions about demandingness and when it might be permissible to preserve one’s own life. [↑](#footnote-ref-92)
93. Colburn (2008, p.102) is quite right to hold that “one must be able to make voluntary choices if one is to live an autonomous life”. In Chapter 5 I will develop a theory of autonomy based on the idea of responding to reasons by committing to a hierarchy of principles and thereby forming a distinct identity. Voluntariness is an important element of that story. [↑](#footnote-ref-93)
94. External sanctions are, of course, often used to cultivate internal sanctions but this possibility does not affect the question of appropriateness. For some discussion on this point see Korsgaard (1996a, p.79-80; p.84) on Mill. [↑](#footnote-ref-94)
95. Korsgaard (1996a, p.88) makes a distinction between demands of morality and the demands of moral feelings but, obviously, in an ideal world the demands of the moral emotions would be tied to the demands of morality. [↑](#footnote-ref-95)
96. See Section 2.5. [↑](#footnote-ref-96)
97. See Section 2.6. [↑](#footnote-ref-97)
98. Korsgaard also uses the relationship between perception and belief as an example but for ease of exposition I will stick to desires and actions here. [↑](#footnote-ref-98)
99. “I ought never to act except in such a way that I could also will that my maxim should become a universal law” (Kant 1998, p.15/4:402). For an interesting discussion about how the Formula of Universal Law fits with the other formulations see also Korsgaard (1996b, Ch.3). [↑](#footnote-ref-99)
100. Hegel’s empty formalism objection, for example, is based on the difficulties involved in making this step. See Knowles (2002, pp.198-210) for a helpful discussion. [↑](#footnote-ref-100)
101. Korsgaard bypasses the second formulation here, the Formula of Humanity, in favour of the Formula of the Kingdom of Ends. I suspect this is because as a member of the kingdom of ends one endorses a more particular identity as a part of special kind of community. Nothing much hangs on her choice, it just makes her exposition a little easier. This is not to say that she does not think the Formula of Humanity is important, in fact quite the opposite. However, the later formulation supports a more concrete identity. [↑](#footnote-ref-101)
102. I think the best way to explain this is, again, by using Raz’s exclusionary reasons. That a die-hard fan will be late for kick-off is excluded as a relevant consideration when there is a child drowning nearby. However, a fan’s moral reasons are not silenced by the demands of following a team. So, someone can care an awful lot about football and not all that much about morality, but if she still identifies as a moral person then this identity will play a regulative role. See Raz (1990, p.38) for his own exposition of the idea of exclusionary reasons. [↑](#footnote-ref-102)
103. See Williams (1999, Ch.1) on character and morality. [↑](#footnote-ref-103)
104. Section 2.1. [↑](#footnote-ref-104)
105. That is to say, she holds that the value of humanity in our own persons is inescapable and so the reflective structure of our own consciousness guarantees that we identify as citizens of the kingdom of ends and are thus bound. [↑](#footnote-ref-105)
106. “Yet most of the self-conceptions which govern us are contingent” (Korsgaard 1996a, p.120). [↑](#footnote-ref-106)
107. This is similar in structure, but ultimately different, to the ‘wrong sort of reasons problem’ in value theory because the issue here is not what makes an option valuable, but what makes it obligatory. See Hieronymi (2005) for a useful discussion of the problem. [↑](#footnote-ref-107)
108. It is not impossible to imagine a character who does not care about such things as having plans or organising her life around certain principles, projects, and ends. However, if such a person did exist, she would unreliable in the extreme and, it is worth pointing out, would be unable to form the kinds of lasting interpersonal relationships that contribute so profoundly to the value of most of our lives. [↑](#footnote-ref-108)
109. See Section 2.6. [↑](#footnote-ref-109)
110. Nagel is concerned that on an identity-based sanction view, self-interest is required to determine the content of our commitments. As we shall see he is mistaken because self-interest is instead securing the appropriate relationship between an agent and her commitments. [↑](#footnote-ref-110)
111. See Nagel (1986) and (1991) for his developed position. [↑](#footnote-ref-111)
112. This is similar to Anscombe’s (1958) worry about the absurdity of self-legislation that I discussed in the last chapter. However, Cohen is not so much concerned with the structure or the bindingness of moral obligations. Rather, his worry is chiefly about malleability of our interests. We do have some control over them and, to an extent, can decide what we care about. [↑](#footnote-ref-112)
113. Which is why my view does not fall prey to Nagel’s (1996) objection discussed in Section 3.2 above. [↑](#footnote-ref-113)
114. My view owes much to Kant but is not a Kantian view. For the perils of carelessly appropriating Kant’s ideas and legacy see O’Neill (2003a). [↑](#footnote-ref-114)
115. This is not to say that there should be no place for spontaneity or whimsicality. These are important parts of a flourishing human life. There are limits though, and it is important to balance our need to be impulsive once in a while with the other elements of a good life. For a useful guide see Nussbaum’s (2000, pp.78-80) well-known list of the central capabilities, which includes play. [↑](#footnote-ref-115)
116. See Darwall (1985, esp. pp.101-113) for a discussion of the unity of agency. “Were we but a bundle of such individual preferences we would have no way of coming to and expressing one mind on the question of what to do; for we would have no perspective other than that internal to each individual preference from which to order our different individual preferences, consider how to deal with conflicts between them, and decide what *we* prefer, on the whole, to do” (Darwall 1985, p.103). [↑](#footnote-ref-116)
117. Section 2.2. [↑](#footnote-ref-117)
118. While their respective behaviour might provide such a reason. [↑](#footnote-ref-118)
119. “Rationality consists, at least partly, in our capacity to make our ends and preferences the object of our rational consideration and to revise them in accordance with reasons we find compelling” (Darwall 1985, pp.101-102). [↑](#footnote-ref-119)
120. Of course, this also serves communities by ensuring that children grow up to be functional adults. See Marcuse (2008) for a discussion of authority, freedom, and socially-necessary repression. [↑](#footnote-ref-120)
121. See Bratman (2001; 2007, esp. Ch.11) for an interesting discussion about agential authority and how some principles can play a governing psychological role. For how we can decide to reaffirm values to which we are already committed see Colburn (2010, p.23). [↑](#footnote-ref-121)
122. Well, safer. Rocky could also learn how to absorb punches with some part of his body that isn’t his head. [↑](#footnote-ref-122)
123. This is a term that Scanlon (1998) introduces to explain his thought that desires are really a particular species of reason. I do not want to go anywhere near that territory here, but the term perfectly captures what I want to communicate so permit me a very brief digression. Scanlon (1998, p.39) says: “A person has a desire in the directed-attention sense that P if the thought of P keeps occurring to him in a favorable light, that is to say, if the person’s attention is directed insistently toward considerations that present themselves as counting in favour of P”. A desire, for Scanlon, is really just the feeling of being preoccupied with certain reasons that count in favour of a particular action. The pleasing taste of a cold beer on a sunny day for example, or how it would relax you from the mental stress of philosophising. [↑](#footnote-ref-123)
124. I will expand on this thought in Section 5.5. [↑](#footnote-ref-124)
125. For a discussion of moral obligation and Kant’s conception of the holy will see Stern (2013). [↑](#footnote-ref-125)
126. See the poem *Kind of an Ode to Duty* by Ogden Nash. [↑](#footnote-ref-126)
127. “It is inconceivable that reason can reject as beneath it emotions to which the heart assents with gladness, and if humans, however morally abject, are not incapable of rising at least in their own estimation” (Schiller 2005, p.152). [↑](#footnote-ref-127)
128. At least, we can endorse them voluntarily. Once we do it is unacceptable to violate them, but we can embrace them in the first place because of the appeal of the justification behind them. See Section 2.7. [↑](#footnote-ref-128)
129. Most of the time we are doing well to live with what Schiller calls dignity. “Only by crushing the power of desire, which rushes too eagerly towards satisfaction, and would prefer to skirt around the will’s authority altogether, do human beings display their independence and prove themselves to be moral beings, which never simply desire or simply loathe, but have to will their loathing and desire in each instance” (Schiller 2005, p.157). [↑](#footnote-ref-129)
130. Schiller himself (2005, p.132) emphasises the operation of the will and the capacity of human beings to “change in accordance with reasons”, so grace is not entirely accidental. However, it is not entirely deliberate either since “[a]s soon as we notice that the *grace* is artificial, our hearts close, and our souls, which surged upwards, now flee backwards. Suddenly, we see that mind has become matter and heavenly Juno has become a cloud” (Schiller 2005, p.139, original emphasis). Grace, for Schiller, is a matter of expressing one’s freedom by willing the moral law without resistance from nature and so it cannot be forcefully achieved by artifice. [↑](#footnote-ref-130)
131. Habermas, of course, ultimately comes to hold a more complicated view of legitimacy, along the lines of the sufficiency thesis, which relies not on the beliefs that people have about the legitimacy of authority and power, but rather on the possible justification of such beliefs. See Habermas (1979; 1997). Knight and Johnson (1994, p.284) defend a similar view, and believe that “a political outcome is legitimate, if at all, because it survives the deliberative process, because it is produced by the sort of reasoned argumentation under fair procedures that defines deliberation as a critical ideal”. [↑](#footnote-ref-131)
132. Buchanan (2002, pp.689) also thinks that a normative account of legitimacy is required in order to tell “what beliefs in legitimacy are beliefs about”. [↑](#footnote-ref-132)
133. For a discussion of how a natural duty of justice might require us to support just institutions in order to secure their effectiveness see Waldron (1993, esp. pp.20-27). For Waldron, effectiveness is a multi-faceted notion of which popular support is only one, albeit important, element. In cases where there is more than one prospective authority passing the test of justice and claiming the right to rule, we must establish which one we should favour with our support. A greater likelihood of one side securing a monopoly over the use of force is one consideration that we will have to take into account. [↑](#footnote-ref-133)
134. Freeman (1990, p.122) argues that this understanding of the basis of legitimacy “is based on the liberal idea that the legitimacy of social rules and institutions depends on their being freely and publicly acceptable to all individuals bound by them”. While my discussion certainly fits with a liberal interpretation, I hope it will have broader relevance. [↑](#footnote-ref-134)
135. For further discussion of the relationship between legal and other types of obligation see Hart (1994, esp. pp.167-179). [↑](#footnote-ref-135)
136. The corollary of this is that it appears to be possible to pull our three varieties of legitimacy apart in a number of different ways. Can we, for instance, have a case of legal legitimacy without meeting our criteria for political or moral legitimacy? (Knowles 2010, p.23) The short answer is yes, but the same direction of explanation persists. The issue of legal legitimacy can only meaningfully be settled within a context of assumed political legitimacy, which in turn yields to a more demanding standard when we want to get to grips with the moral situation. Explaining to a revolutionary that she has broken a law and should suffer the consequences outlined in the statute books will be met with incredulity because she is challenging the very status of those statutes. [↑](#footnote-ref-136)
137. We can break this down further by discussing cases of governments or subordinate institutions acting *ultra vires*, or instances where states violate international laws or treaties, and so on. For a helpful discussion see Knowles (2010, p.20-28). [↑](#footnote-ref-137)
138. Hobbes (1985, p.375) likens citizens who feel free to dispute the Sovereign’s power to parasitic roundworms, a nuisance that is difficult to cure and potentially fatal. See Hobbes (1985, esp. Ch.29) for a general discussion of threats that can weaken or even dissolve a commonwealth. [↑](#footnote-ref-138)
139. See Berlin (1981) for a comprehensive taxonomy of the perspectives that one can take on Machiavelli’s philosophy. [↑](#footnote-ref-139)
140. Section 4.2. [↑](#footnote-ref-140)
141. See Sections 2.6, 3.4, and 3.5. [↑](#footnote-ref-141)
142. See Section 1.5 for the distinction between weak and strong philosophical anarchism. [↑](#footnote-ref-142)
143. See Buchanan (2002, pp.703-709) and also Christiano (2004, pp.105-131). [↑](#footnote-ref-143)
144. I develop this idea in Chapter 6. [↑](#footnote-ref-144)
145. See Rawls (1999, p.99). [↑](#footnote-ref-145)
146. For simplicity’s sake assume that this qualification follows all occurrences of ‘just state’ unless otherwise indicated. See Section 1.3 where I discuss Waldron’s explanation of how this is a range-limited principle. [↑](#footnote-ref-146)
147. See Sections 2.4 and 3.4 for a more detailed explication of Raz’s theory of authority and how I have adapted it here. [↑](#footnote-ref-147)
148. McTernan has described how experiments about the effect of social norms show that individuals alter their behaviour when presented with a new rule that they take to be obligatory. For example, telling hotel guests that reusing towels saved the environment and that other guests reused their towels increased the numbers that did so. See McTernan (2014, pp.94-100) [↑](#footnote-ref-148)
149. Whether or not she actually does extract the cost is only of secondary importance. What matters is that it if she did attend to the mismatch between her action and her principles she would find that her identity has been compromised. [↑](#footnote-ref-149)
150. He concludes (2002, p.693) that “neither Rawls nor Raz find the question of political authority (as including the right to be obeyed) as of much consequence for political theory”. [↑](#footnote-ref-150)
151. See Section 1.5 for a discussion, and rejection, of Ladenson’s (1980) view. [↑](#footnote-ref-151)
152. In the sense that obligations are limited in scope. For example, if I sign up to the army then I am bound to obey my drill instructor when he tells me to get down in the mud and give him fifty. I am not required to obey him if he tells me where to spend my summer holidays. [↑](#footnote-ref-152)
153. “I tend to think that in contemporary usage, ‘duty’ does not signify a normatively distinctive category, other than the fact that only categorical reasons, that is, ones whose application is not conditional on the agent’s inclinations or preferences, and so on, can give rise to duties” (Raz 2010, pp.290-291). Buchanan’s view is very similar to Darwall’s critique of Raz on this point. See Darwall (2010) and Raz (2010) for their back and forth. [↑](#footnote-ref-153)
154. My dependents are another story. [↑](#footnote-ref-154)
155. Another possible objection here is that if legitimacy is tied to obligation then states cannot have legitimacy since we cannot owe obligations to abstract, impersonal entities. One way around this is to focus on governments rather than states, where ‘government’ denotes the group of actual individuals who occupy the political offices of a state. Another solution is to allow concrete individuals to transfer some of their rights to the state, thereby empowering it to act as their agent in certain circumstances. Ultimately, however, this objection is a distraction from the main point which is whether or not we can relate to our reasons to obey a suitably just state in an obligatory fashion. As such, I will set it aside here. [↑](#footnote-ref-155)
156. It is worth pointing out here that we can have duties without a corresponding right. These are traditionally known as the imperfect duties. The duty of charity is the best example. I do not owe it to any particular individual to make some charitable contribution but it is plausibly a duty nonetheless. I consider obligatory ends in Section 6.3. [↑](#footnote-ref-156)
157. For a ‘samaritan’ account of political obligation as a response to dire need see Wellman (1996; 2005). [↑](#footnote-ref-157)
158. He calls this the Robust Natural Duty of Justice. See Buchanan (2002, pp.703-709). [↑](#footnote-ref-158)
159. See Section 4.8. [↑](#footnote-ref-159)
160. Brettschneider (2012, pp.38-39) suggests that there is some special democratic value missing in such cases. I find this line of thought difficult to follow. In Chapter 5 I explain the value for individuals in relating to their institutions in an autonomous way. [↑](#footnote-ref-160)
161. See Sandel (1984) for the importance of recognising the ‘encumbered self’. [↑](#footnote-ref-161)
162. See Section 2.7 for a discussion of Olsaretti’s conception of voluntariness. [↑](#footnote-ref-162)
163. A capacity for a sense of justice and a capacity for a conception of the good. See Rawls (1999, p.442). [↑](#footnote-ref-163)
164. See Brettschneider (2012, pp.47-48) for a helpful discussion of what makes for a hateful viewpoint. [↑](#footnote-ref-164)
165. Brettschneider (2012, p.46) points to Martin Luther King Day and Black History Month in the United States. [↑](#footnote-ref-165)
166. Adults should not be forced to attend political philosophy courses, but the credentials of a state can certainly be enhanced by making civic education available to any and all who can be interested. For more in-depth treatment of civic education see Gutmann (1999) and Callan (1997). [↑](#footnote-ref-166)
167. Brettschneider (2012, pp.86-90) is sensitive to this point. [↑](#footnote-ref-167)
168. See Tsai (2014) on how even the offering of reasons can sometimes constitute overbearing intrusion insofar as it prevents an agent from working through something on her own. [↑](#footnote-ref-168)
169. See Dworkin (1988, pp.5-6) for a classic taxonomy. See also Schneewind (1998) for an incredibly detailed reconstruction of the history of the concept. [↑](#footnote-ref-169)
170. This is not intended as an exhaustive categorisation of the values expressed in debates about autonomy. For the purposes of this chapter I hope that a rough sketch of the conceptual terrain will be sufficient. [↑](#footnote-ref-170)
171. Neuhouser (2000, p.24) identifies this kind of autonomy with personal freedom in the Hegelian schema: “the will’s choosing of its own ends”. [↑](#footnote-ref-171)
172. Oshana (1998, p.82) describes these goals as “relevant to the direction of her life” and argues that the agent must be able to “pursue these goals and make them effective in action”. [↑](#footnote-ref-172)
173. “When we say that a person is self-governing because she is in control of her actions and choices, we are saying more than that the person’s actions coincide with preferences or values that are her own. We are saying that the person has the power to determine how she shall live” (Oshana 1998, p.82). [↑](#footnote-ref-173)
174. As Feinberg (1980, p.21) says: “I am autonomous if I rule me, and no one else rules I.” [↑](#footnote-ref-174)
175. Providing a robust sphere of personal autonomy is likely to require widespread dispersal of what Pettit (1996) calls ‘anti-power’. This removes the possibility that an agent’s personal autonomy is contingent on the continued benevolence of another person. [↑](#footnote-ref-175)
176. See Christman (1991a, p.17) on this point: “Delusion, paranoia, and other psychopathologies will be inconsistent with autonomy because of the suffering agent’s inability to make consistent and reflective judgments about her own set of desires by which she is moved to action”. [↑](#footnote-ref-176)
177. Mill’s harm principle is the stock example of how we might constrain the sphere of personal autonomy so that we all enjoy a similar level of control. [↑](#footnote-ref-177)
178. Of course, it may be the case that someone holds a conception of the good that specifies having no control whatsoever over their actions or environment but in the absence of at least some domain of negative freedom this is not something that can be *pursued*. It is also worth drawing a distinction between pursuing some particular good and pursuing some conception of the good. A measure of personal autonomy is a prerequisite for both. However, a conception of the good is necessary to determine whether a particular good is worth chasing and in under what circumstances. [↑](#footnote-ref-178)
179. See Section 4.5. [↑](#footnote-ref-179)
180. See Hegel (2010, esp. Part 2 & Part 3 Sec. 3). [↑](#footnote-ref-180)
181. See Section 4.6. [↑](#footnote-ref-181)
182. McDowell (1996, p.9) attributes this view to Kant: “When Kant describes the understanding as a faculty of spontaneity, that reflects his view of the relation between reasons and freedom: rational necessitation is not just compatible with freedom but constitutive of it. In a slogan, the space of reasons is the realm of freedom”. McDowell’s own view of the relation between reasons and freedom is similar. He holds that free agency is made possible by second nature, i.e. the natural operation of our cultivated rational faculties in conjunction with the world as it presents itself to us. I suspect that there is interesting common ground to be explored between McDowell’s notion of second nature and my adaptation of Korsgaard’s idea of practical identity, but I will leave that project aside for another time. [↑](#footnote-ref-182)
183. Hurka (2011, p.149) continues: “Then [an agent’s] intentions are hierarchically arranged, and exercise sophisticated rationality”. [↑](#footnote-ref-183)
184. See Neuhouser (2000, especially pp.82-114) for an account of the subjective component of social freedom in Hegel’s philosophy. Although I draw on several elements of the Hegelian picture, I think the view I offer here is distinct at least insofar as it is a much, much more limited account of the interest that creatures such as ourselves have in responding to reasons in various ways. [↑](#footnote-ref-184)
185. For additional reasons for pessimism about the possibility of using ‘autonomy’ as something more than a term of art see O’Neill (2003a) and Dworkin (1988). See Gallie (1955) for the idea that concepts can be “essentially contested”. [↑](#footnote-ref-185)
186. Hurka (2011, p.142) also suggests that being a slave is not necessarily a barrier to exercising one’s rational capacities since one might still be capable of internally evaluating possible actions and states of affairs. [↑](#footnote-ref-186)
187. Schier’s target here is MacCallum (1967) and his tripartite definition of freedom. MacCallum could conceivably respond that it is our freedom to enjoy control over our body or our access to sensory information that is at stake and so his definition emerges unscathed. However, the flexibility of MacCallum’s account is also its major weakness. Schier is right to argue that the appeal of certain kinds of freedoms are more comfortably accommodated as either positive or negative freedoms, showing that MacCallum jettisons a useful, if problematic, distinction too quickly. [↑](#footnote-ref-187)
188. It is also worth noting that: “Although negative freedom is an intrinsic value, particular negative freedoms are not unconditional values, for their intrinsic value is a function of the other things we value” (Schier 1993, p.15). [↑](#footnote-ref-188)
189. See Raz (1986, p.410) for a different instrumental view of negative freedom. [↑](#footnote-ref-189)
190. Hayek (1991), for example, places great emphasis on the importance of the predictability of the state. [↑](#footnote-ref-190)
191. As I understand him, Schier draws no hard line between physical space and social space. Since we exist in a cultural as well as a physical landscape, the same basic interest can explain the demand for freedom of expression as explains the demand for bodily integrity. [↑](#footnote-ref-191)
192. As Tiberius says: “it is as *reflective* creatures that we want to know how to live… To abandon the reflective self in our account of how to live would be to ignore the real source of our questions about how to live and, hence, to risk not really answering the questions we have” (Tiberius 2008 p.5, original emphasis). [↑](#footnote-ref-192)
193. We must be sensitive to contextual factors here. Considerations can count in favour of different actions in different circumstances and some features of a situation that might otherwise be salient could be motivationally inert. [↑](#footnote-ref-193)
194. In some more or less carefully defined sphere. See Section 2.4. [↑](#footnote-ref-194)
195. While there are religious and cultural contexts in which it is instrumentally valuable for individuals to conform to formal or informal codes of grooming and appearance, participation by individuals is generally supposed to meet a standard of voluntariness. The choice of an individual to become a Tibetan monk, for instance, or a goth, is a response to reasons on their part and, therefore, forms part of their attempt to lead an intelligible life. [↑](#footnote-ref-195)
196. See Raz (1986, pp.308-13) for an argument to the effect that a person can “only have a comprehensive goal if is based on existing social forms, i.e. on forms of behaviour which are in fact widely practised in his society” (Raz, 1986, p.308). [↑](#footnote-ref-196)
197. It will, of course, be much worse for anyone who is targeted by such a law. [↑](#footnote-ref-197)
198. See Aristotle (1992, pp.53-62/1252a1-1253a29). [↑](#footnote-ref-198)
199. See Section 2.6. [↑](#footnote-ref-199)
200. See Section 1.5. [↑](#footnote-ref-200)
201. See Locke (2004, pp.269-278) for his conception of the state of nature. [↑](#footnote-ref-201)
202. See Kant (2008, p.52/4:446-447) for a classic statement of the contrast. [↑](#footnote-ref-202)
203. See Section 3.2. [↑](#footnote-ref-203)
204. I believe Aristotle (2009, p.11/1098a) may have had something like this in mind when he distinguished between the capacity to be obedient to reason and the capacity to exercise thought: “There remains, then, an active life of the element that has reason; of this, one part has it in the sense of being obedient to reason, the other in the sense of possessing reason and exercising thought”. [↑](#footnote-ref-204)
205. And who is not trying to be extra clever and surprise the critics with a conventional piece! [↑](#footnote-ref-205)
206. See Section 2.7. [↑](#footnote-ref-206)
207. Or some desire, decision, choice, etc… I prefer to keep the focus on a person rather than on a property of a person. [↑](#footnote-ref-207)
208. I will not even attempt to list every procedural account of autonomy here. The three I have referenced are, however, very broadly representative of a wide range of procedural views. [↑](#footnote-ref-208)
209. See Section 5.4. [↑](#footnote-ref-209)
210. “Man was born free, and he is everywhere in chains. Those who think themselves the masters of others are indeed greater slaves than they. How did this transformation come about? I do not know. How can it be made legitimate? That question I believe I can answer” (Rousseau 1968, p.49). [↑](#footnote-ref-210)
211. For my understanding of unintelligibility see Sections 2.6 and 3.4. [↑](#footnote-ref-211)
212. See Section 6.6 where I will disambiguate some of the concerns that often get lumped together in debates about ideal and non-ideal theory. For some helpful ‘conceptual cartography’ see Valentini (2012). [↑](#footnote-ref-212)
213. See also Gauthier (1986, p.339). [↑](#footnote-ref-213)
214. “It is because contractarians’ conceptions of the person differ that their descriptions the state of nature differ, and this partially explains why the reasons they invoke to justify the state’s creation differ, and why the structures of the states they advocate differ” (Hampton 1986, p.273). [↑](#footnote-ref-214)
215. See Plato (2005, p.615/368e). [↑](#footnote-ref-215)
216. See Plato (2005, Books VIII and IX). [↑](#footnote-ref-216)
217. This may also explain the expulsion of the mimetic artists from the republic. Art invites us to identify with its subjects, and if those subjects strongly exhibit tendencies that would undermine one’s self-image as a conscientious citizen then that process of identification might lead to a change in one’s own commitments. Plato’s fears in this regard are due to the limitations of the rigid tripartite structure he attributes to the soul. [↑](#footnote-ref-217)
218. For example, ambition or acquisitiveness. [↑](#footnote-ref-218)
219. Reason, spirit, or appetite on Plato’s tripartite theory. He is at least right in thinking that our various concerns and interests attract us to more than one practical identity. [↑](#footnote-ref-219)
220. See Chapters 2, 3, and 5. [↑](#footnote-ref-220)
221. See Section 4.4. [↑](#footnote-ref-221)
222. This is certainly an accurate reading of Aristotle (1992, p.61/1253a18): “It is clear then that the state is both natural and prior to the individual”. Although, he also notes (1992, p.187-188/1278b15-37) that the purpose of a state cannot be understand without understanding the good of its members. See also Aristotle (1992, p.168/1274b32). It is noteworthy that Hampton points to Socrates here and not Plato. Presumably she is thinking of the earlier dialogues, but it is an interesting question nonetheless whether, and if so why, she holds that Plato came to prioritise the individual in some sense. [↑](#footnote-ref-222)
223. For a classic statement see Cohen (1989). [↑](#footnote-ref-223)
224. The original position “models what we regard – here and now – as acceptable restrictions on the reasons on the basis of which the parties (as citizens’ representatives), situated in those fair conditions, may properly put forward certain principles of justice and reject others” (Rawls 2003, p.80). [↑](#footnote-ref-224)
225. Rawls (1999, p.413) grants the importance of role models in the context of his three stages of moral psychology in *A Theory of Justice*: “But we may also suppose that the newer members of the association recognize moral exemplars, that is, persons who are in various ways admired and who exhibit to a high degree the ideal corresponding to their position. These individuals display skills and abilities, and virtues of character and temperament, that attract our fancy and arouse in us the desire that we should be like them, and able to do the same things… they exhibit qualities and ways of doing things that appeal to us and arouse the desire to model ourselves after them”. However, he does not seem to realise that, in many ways, those of us to whom his principles of justice are really addressed could use role models of our own. [↑](#footnote-ref-225)
226. See Kant (1996, p.32/6:240). [↑](#footnote-ref-226)
227. See Rawls (1999, p.99). [↑](#footnote-ref-227)
228. This perhaps overstates things a little. It will, for example, be quite difficult to determine exactly when parents stop sending their children to school if they allow them to miss several days every week, but this is a general difficulty with fixing the performance conditions of an action. [↑](#footnote-ref-228)
229. See Baron for helpful discussion on this point, especially (1995, pp.30-31). [↑](#footnote-ref-229)
230. See Section 5.5. [↑](#footnote-ref-230)
231. See Kant (1996, pp.201-203/6:452-6:454) on the duty of beneficence. See also Kant (1998, p.33/4:423). [↑](#footnote-ref-231)
232. It is worth pointing out that a conception of the person is really more implicit in classical theorists such as Hobbes (1985) and Locke (2004). [↑](#footnote-ref-232)
233. “As such, the parties are artificial persons, merely inhabitants of our device of representation: they are characters who have a part in the play of our thought experiment” (Rawls 2003, p.83). See also Rawls (2005, p.xxix; pp.22-28). [↑](#footnote-ref-233)
234. “Whether these two points of view are congruent is likely to be a crucial factor in determining stability” (Rawls 1999, p.497). [↑](#footnote-ref-234)
235. I shall start by looking at Rawls’s idea of stability as it figures in *A Theory of Justice* and later consider how his thinking on the subject changes in *Political Liberalism*. See Barry (1995) for a painstaking discussion and critique of the transition. [↑](#footnote-ref-235)
236. See also Rawls (2005, p.19). [↑](#footnote-ref-236)
237. Elsewhere he emphasises the regulative role that a sense of justice plays and how it binds individuals to acknowledge reasons of justice: “a just person is not prepared to do certain things, and so in evil circumstances he may decide to chance death rather than act unjustly… the just man does what all things considered he most wants” (Rawls 1999, p.502). [↑](#footnote-ref-237)
238. Until now, critical attention has been focused on the necessity, and indeed the viability, of an overlapping consensus between radically divergent worldviews. I am sympathetic to the arguments of Barry (1995) and Baier (1989), who contend that an overlapping consensus is redundant in the Rawlsian framework. I will suggest, however, that the overlapping consensus that Rawls pursues is important when we see the value of inverting the traditional contract architecture. I am interested here in the way that Rawls conceives the question of stability in the first place. [↑](#footnote-ref-238)
239. It is important to note that this is already an additional axis of evaluation for contract theories, along with coordination and efficiency, although it is subordinated to considerations of right. The stability of a conception of justice is important if it is to be deployed to realise justice, but it is not internal to the question of which principles are just. ““We cannot, in general, assess a conception of justice by its distributive role alone, however useful this role may be in identifying the concept of justice. We must take into account its wider connections; for even though justice has a certain priority, being the most important virtue of institutions, it is still true that, other things being equal, one conception of justice is preferable to another when its broader consequences are more desirable” (Rawls 1999, pp.5-6). [↑](#footnote-ref-239)
240. See Rawls (1999, p.349). [↑](#footnote-ref-240)
241. For an exposition of the Aristotelian principle see Rawls (1999, p.500). [↑](#footnote-ref-241)
242. “One may say, then, that a person who lacks a sense of justice, and who would never act as justice requires except as self-interest and expediency prompt, not only is without ties of friendship, affection, and mutual trust, but is incapable of experiencing resentment and indignation. … Put another way, one who lacks a sense of justice lacks certain fundamental attitudes and capacities included under the notion of humanity. … This liability is the price of love and trust, of friendship and affection, and of a devotion to institutions and traditions from which we have benefited and which serve the general interests of mankind” (Rawls 1999, p. 428) [↑](#footnote-ref-242)
243. As defined by the thin theory of the good. See Rawls (1999, pp.347-350). [↑](#footnote-ref-243)
244. The deliciousness of a piece of cake for example. I know it is bad for me, and I know I will regret eating it. However, if I continue to focus on how delicious it is then, other things being equal, sooner or later I’m going to scoff it down. [↑](#footnote-ref-244)
245. See Chapters 2 and 3. [↑](#footnote-ref-245)
246. There are connections here, I think, with the account of grace that I sketched in Section 3.6. [↑](#footnote-ref-246)
247. See Rawls (2005, pp.xv-xvi). [↑](#footnote-ref-247)
248. Reasonable people “desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept. They insist that reciprocity should hold within that world so that each benefits along with others” (Rawls 2005, p.50). [↑](#footnote-ref-248)
249. “But for one who understands and accepts the contract doctrine, the sentiment of justice is not a different desire from that to act on principles that rational individuals would consent to in an initial situation which gives everyone equal representation as a moral person. Nor is it different to from wanting to act in accordance with principles that express men’s nature as free and equal rational beings” (Rawls 1999, p.418). [↑](#footnote-ref-249)
250. “The problem of political liberalism is to work out a conception of political justice for a constitutional democratic regime that the plurality of reasonable doctrines – always a feature of the culture of a free democratic regime – *might* endorse” (Rawls 2005, p.xvii, my emphasis). See also: “what counts is the kind of stability, the nature of the forces that secure it… The kind of stability required of justice as fairness is based, then, on its being a liberal political view*, one that aims to be acceptable* to citizens as reasonable and rational, as well as free and equal, and *so as addressed to their public reason*” (Rawls 2003, p.185, my emphasis). And: “as a liberal political conception, justice as fairness is not reasonable in the first place unless it generates its own support in a suitable way *by addressing each citizen’s reason*, as explained within its own reason” (Rawls 2003, p.186, my emphasis). [↑](#footnote-ref-250)
251. See Rawls (1999, pp.7-8). [↑](#footnote-ref-251)
252. Although it does not directly address the ideal/non-ideal divide see Caney (2014) for an interesting discussion of the climate preservation duties that accrue to conscientious individuals precisely because many people fail to do their fair share to avert disaster. See also Miller (2011). [↑](#footnote-ref-252)
253. See Farrelly (2007). [↑](#footnote-ref-253)
254. See Sen (2009) and Simmons (2010). [↑](#footnote-ref-254)
255. See Section 4.8. [↑](#footnote-ref-255)
256. See Section 6.3. [↑](#footnote-ref-256)
257. “Adversity, pain, and want are great temptations to violate one’s duty” (Kant 1996, p.152/6:388). [↑](#footnote-ref-257)
258. It is worth noting that Rawls is often attacked from the other side for being *too* concerned with incorporating constraints like moderate scarcity to model the real world. For such a position see Cohen (2003; 2008). Valentini (2012) suggests that this criticism is motivated less by methodological questions arising from the question of what we should do to make the world more just than it is by questions about what we should think about justice and what it means. I do not believe there is a very sharp distinction here but that debate will have to wait for another day. [↑](#footnote-ref-258)