

# Liberal Theories of Political Authority

by

Steven Coyne

A thesis submitted in conformity with the requirements  
for the degree of Doctor of Philosophy

Department of Philosophy  
University of Toronto

© Copyright 2021 by Steven Coyne

# Liberal Theories of Political Authority

Steven Coyne

Doctor in Philosophy

Department of Philosophy  
University of Toronto

2021

## Abstract

In this thesis, I argue that liberal theory faces a fundamental problem in accounting for the legitimate authority of the state. On the one hand, legitimate authority entails that the state has a right to the obedience of its citizens – a right that citizens not act on some of the considerations that they find morally important, and that they instead simply do what the state tells them to do. On the other hand, liberalism is committed to respect for autonomy, which entails that citizens have a moral obligation to act on their moral reasons as they understand them. To reconcile the two, liberalism must make sense of how a citizen could see themselves as required to refrain from acting on considerations that they believe to be morally relevant, and instead act on what the law demands of them. I frame a number of the leading theories of authority as attempts to explain how this is possible, including Rawls’s justice as fairness, Raz’s service conception of authority, and Darwall’s second-person standpoint. Ultimately, however, none of these theories appear to satisfactorily resolve the problem. This suggests that liberals may have to accept that their states lack legitimate authority over them.

# Acknowledgments

First of all, I owe a great deal to my committee.

David Dyzenhaus supervised this project from beginning to end. He knows more than anyone the extent to which this thesis is the tip of an iceberg. He read and commented on countless drafts of chapters that did not work, and many versions of the ones that did, always within days, often within hours. Although his mark is felt on every page of this thesis, I think that most of all it is felt in its overall vision. To the extent that this thesis can speak to problems in political philosophy and political theory of any enduring significance, and to the extent that it tells a unified story, that is the result of David's mentorship.

Arthur Ripstein influenced this thesis in crucial ways. Some of the comments I received from Arthur were essentially essays onto themselves, somehow sandwiched into a single Microsoft Word comment box. The best-developed parts of my argument invariably bear the influence of those comments.

Joseph Heath was also an important mentor to me. He was very generous with his time, and he helped me a lot when I was orienting myself in political philosophy, of which I had little knowledge before writing this thesis. I also learnt a lot from his general acumen as a teacher and as a writer. A substantial chunk of what I learnt about writing in grad school, I learnt from reading his posts on *In Due Course*, which I really hope will be resurrected someday.

Thank you also to Gurpreet Rattan and Kenneth Ehrenberg, who served as internal and external examiners for this thesis, for their thoughtful questions. Although I unfortunately did not have enough time to incorporate their substantive comments into the final version of this thesis, the impact of those comments will be felt down the line. In the shorter term, I am grateful that they caught many readily fixable errors that would have gnawed at me for years to come.

The best part of graduate school was the people whom I've befriended: Mark Fortney, Valerie Bernard, Benjamin Wald, Larkin Philpot, Emma McClure, Kevin Kuhl, Michael Da Silva, Maggie O'Brien, Katie Fitzpatrick, Robyn Forman, Jacob Stump, Hillary Nye, Gillman Payette, Eric Shoemaker, Johanna Thoma, Parisa Moosavi, Elena Rabinoff, Dan Rabinoff, Mason Westfall, Michaela Manson, Melissa Rees, Hamish Russell, Belinda Piercy, C Dalrymple-Fraser, Dominic Alford-Duguid, Luke Roelofs, Harry Krahn. I'll add the department fantasy football team, the Tranzac, the former Green Room, the former Sakura Sushi, and the former Victory Cafe as honorary

non-human members of this list. All of you have made me a better philosopher, and have given me some adventures to remember along the way.

Three of my graduate school friends deserve special acknowledgement. Eric Mathison reintroduced me to competitive running, which kept me physically and spiritually healthy while writing this thesis. The comfort of our friendship was amply demonstrated when we ran a 100-miler together without uttering more than a handful of sentences to one another. Josh Brandt was my roommate and confidant during the most turbulent years of my PhD, and the person with whom I've worked most frequently in teaching. I often find myself asking, when I'm frustrated with some student interaction, or really in any situation that requires empathy, "What would Josh Brandt do?", and I never regret acting on that. Jeremy Davis is 50% responsible for what is probably the longest message chain on my phone. Our conversations about life, sports and philosophical apprenticeship have made me a better philosopher, improved my taste in books, music and movies, and deepened my overall humanism.

My friendships from outside of academic philosophy also sustained me through the writing process. In particular, I owe thanks to Bryan Johnston-Lemke, Mary Riehle, Barbara Shain, Josh Zukewich, David Brandman, Sonja Babovic, Andrea Wall, Penelope Gorton, Liv Capozzi, Marcus Brubaker, Megan Greischar, Sarah Hurowitz, Benjamin Riou, and Kelly Marie Paton. My friends in the Toronto running community are too numerous to mention, but they helped me enormously, too. Apologies to anyone whom I've left off this list by accident.

Thank you to my parents, Joyce Steiner and Tim Coyne, for keeping me curious and skeptical as a child, and for encouraging my voracious reading habits. It was inevitable that they would have at least one philosopher as a child, and since my brother Jason counts as an honorary philosopher anyways, it seems that they went two for two. My partner's parents, Jan and Joe, welcomed me into their home for much of the final push of writing this thesis – thank you.

My most significant thank you goes to Nicole Mideo. You have supported me at every step of this journey, most of all through this madcap past year, where you took so many burdens off of my shoulders, even though you already had so many of them on yours. You bring intelligence, fastidiousness, and concern for others to our table, day in and day out. I have learnt so much from you about how to be an effective and good human being.

# Table of Contents

Acknowledgments.....	iii
Table of Contents .....	v
Chapter 1: Introduction .....	1
1 The Problem of Legitimate Authority .....	1
2 Liberal Theories of Political Authority .....	4
3 Survey of the Thesis .....	7
Chapter 2: Wolff's Challenge to Authority .....	14
1 Introduction .....	14
2 Wolff's View of Autonomy .....	15
3 Wolff's View of Authority.....	18
4 The Primary Conflict Between Authority and Autonomy: Exclusion .....	23
5 The Secondary Conflict Between Authority and Autonomy: Content-Independence.....	25
6 Promises and Decisions .....	29
7 The Liberal Response to Wolff .....	30
8 Conclusion .....	32
Chapter 3: Moralism .....	33
1 Introduction .....	33
2 Moralism about the <i>Prima facie</i> Duty to Obey the Law .....	34
3 Moralism about Political Authority .....	36
4 A Solution to Wolff's Challenge?.....	37
5 The Failure of the Solution .....	39
6 Conclusion .....	41
Chapter 4: Justice as Fairness .....	42
1 Introduction .....	42
2 <i>A Theory of Justice</i> : Overview .....	43

3	<i>A Theory of Justice: Political Authority</i> .....	45
4	<i>A Theory of Justice: A Solution to Wolff’s Challenge?</i> .....	50
5	<i>A Theory of Justice: The Failure of the Solution</i> .....	53
6	<i>Political Liberalism: Overview</i> .....	57
7	<i>Political Liberalism: Political Authority</i> .....	62
8	<i>Political Liberalism: No Solution to Wolff’s Challenge</i> .....	64
9	Conclusion .....	65
	Chapter 5: The Service Conception of Authority .....	67
1	Introduction .....	67
2	Exclusionary Reasons .....	68
3	The Service Conception of Authority: Conceptual Aspects .....	69
4	The Service Conception of Authority: Normative Aspects .....	73
5	A Solution to Wolff’s Challenge?.....	82
6	Conclusion .....	86
	Chapter 6: Darwall’s Objection to the Service Conception.....	87
1	Introduction .....	87
2	The Normal Justification Thesis .....	88
3	The Expertise Counterexample .....	89
4	The Motivation Counterexample .....	93
5	Cases with Moral Dependent Reasons.....	98
6	The Right to Rule.....	100
7	Conclusion .....	102
	Chapter 7: Second-Personal Authority .....	104
1	Introduction .....	104
2	The Second-Person Standpoint .....	106

3	Some Distinctions within the Second-Personal: Bipolar Obligations and Moral Obligations .....	108
4	Second-Personal Authority .....	112
5	Darwall's First Theory of Political Authority .....	116
6	A Solution to Wolff's Challenge? .....	118
7	The Failure of the Solution .....	119
8	Darwall's Second Theory of Political Authority .....	121
9	Conclusion .....	123
	Chapter 8: Conclusion.....	125
	Bibliography .....	131

# Chapter 1

## Introduction

### 1 The Problem of Legitimate Authority

Consider a drill sergeant's command to her subordinate to do a push-up. For her subordinate to take her command in the way that she intends for it to be taken, he must not treat it merely as a reminder that he will face a sanction if he does not comply. Nor can he treat it merely as a reason to do a push-up, to be added to the other reasons for and against doing the push-up. Instead, to fully conform to what is required by the command, he must altogether change the way he deliberates about how to act. He must not act on the other considerations that seem to speak for or against doing push-ups, like how much stronger he would like to become, or even how tired he is at the moment. He must simply do the push-up. He must *obey* the drill sergeant.

In *Essays on Bentham*, H.L.A. Hart describes two key features of this response.<sup>1</sup> The first is *peremptoriness*. The peremptoriness of commands lies in the fact that the commander “characteristically intends his hearer to take the commander’s will instead of his own as a guide to action and so to take it in place of any deliberation or reasoning of his own: the expression of a commander’s will that an act be done is intended to preclude or cut off any independent deliberation by the hearer of the merits pro and con of doing the act.”<sup>2</sup> The drill sergeant does not intend for her subordinate to think about the merits of doing the push-up.

The second is *content-independence*. The content-independence of the drill sergeant’s command “lies in the fact that [she] may issue many different commands to the same or to different people and the actions commanded may have nothing in common, yet in the case of all of them the commander intends [her] expressions of intention to be taken as a reason for doing

---

<sup>1</sup> In this, he follows Hobbes in Chapter XXV of *Leviathan*, who writes that command is “when a man saith do this or do not do this yet without expecting any other reason than the will of him that saith it.” (Hobbes, *Leviathan*, 166)

<sup>2</sup> Hart, *Essays on Bentham*, 253.

them.”<sup>3</sup> Whether the drill sergeant commands one push-up, thirty push-ups, or to clean the latrines, she intends that he will do as she says.

This thesis is about the commands that are issued by legal or political systems, which usually (though not necessarily) take the form of laws.<sup>4</sup> Hart holds that legal systems command their subjects, and that laws are thus intended to be taken as peremptory and content-independent. However, as Joseph Raz observes, it is somewhat implausible to attribute peremptoriness in this strict sense to the law.<sup>5</sup> The problem is that unlike soldiers at boot camp, legal subjects seem to be free to continue to deliberate and think for themselves without contradicting the spirit of the state’s authority. As Raz notes, “Surely what counts, from the point of view of the person in authority, is not what the subject thinks but how he acts. I can do all the law requires of me if my actions comply with it. There is nothing wrong with my considering the merits of the law or of action in accord with it.”<sup>6</sup> However, he agrees that laws possess something not so far off from peremptoriness. Following him, we can call this quality the *exclusionary* nature of law, because laws exclude at least some reasons from action.<sup>7</sup> We can illustrate exclusion with the following simple model of how deliberation leads to action: ordinarily in deciding whether to do an action, an agent will take all of the reasons that speak for or against it, add them up, and then either do the action or not do the action depending on which way the balance of reasons tilts. When a reason is excluded, it is simply not added to the final sum, regardless of how the subject evaluates the weight of that reason. Thus, someone can obey an order while thinking at the very same time that it is mistaken or wrong.

---

<sup>3</sup> Hart, *Essays on Bentham*, 254.

<sup>4</sup> I will frequently say that political authority is expressed through laws, but this may not be the only way that it is expressed. I do not take a stance on which channels count as official, or how the state’s overall authority propagates to the individuals who hold offices in it, like police officers or judges. I take it this question is more closely related to the debate about the nature of law, though a theory of authority will need to explain why only these channels possess legitimate authority. I am also not drawing a sharp distinction between political and legal authority in this thesis.

<sup>5</sup> Raz, *The Morality of Freedom*, 39. He grants that there might be cases of authority (perhaps like the drill sergeant) where the command is offered in a peremptory spirit, but that legal authority is not one of these cases.

<sup>6</sup> Raz, *The Morality of Freedom*, 39. For further discussion, see Shapiro, “Authority,” 406.

<sup>7</sup> Raz, *Practical Reason and Norms*, 184-5.

Next, Hart observes that people respond to commands in many different ways. Sometimes they do something other than the commanded action. Other times, they do what they are told, but out of fear of punishment, or for some other reason altogether. But sometimes, they take them in the spirit they were intended, and they obey the person who commanded it. Insofar as this is the case, we might say, following Robert Paul Wolff, that the commander is a *de facto* authority.<sup>8</sup> At a minimum, their commands are recognized by those whom they command as content-independent, exclusionary reasons, and acted upon as such. For Hart, this recognition amounts to the difference between legal authority and pure force, and is necessary for distinguishing legal systems from simpler, more coercive kinds of social order.<sup>9</sup>

It is one thing to say that people *do* treat authorities in this spirit – this is question of *de facto* authority. It is another to say that people *should* treat authorities in this spirit, as a moral matter or even just a rational matter. This is the question of *legitimate* authority. Note that the question of *de facto* authority is not totally disconnected from morality, because it concerns the moral or rational reasons that people *believe* themselves to have. As Wolff writes, “Even the descriptive sense [*de facto* authority] refers to norms or obligations, of course, but it does so by describing what men believe they ought to do rather than by asserting that they ought to do it.”<sup>10</sup> But again this is different from the question of legitimate authority, which concerns the reasons that people in fact have.

There is little doubt that there are some *de facto* authorities. Many people do, for better or worse, treat laws as content-independent, exclusionary reasons. It is much less clear whether there are any legitimate authorities. Indeed, Wolff famously rejects the existence of legitimate authorities in his book *In Defense of Anarchism*. For Wolff, no *de facto* authority has a right to the obedience of anyone, because such obedience is morally impermissible in the first place. I

---

<sup>8</sup> Wolff refers to *de facto* authorities when he writes that “Descriptively, the state may be defined as a group of persons who are acknowledged to have supreme authority within a territory – acknowledged, that is, by those over whom the authority is asserted.” (Wolff, *In Defense of Anarchism*, 5)

<sup>9</sup> That said, it does not get us all the way to a legal order, which Hart holds to be considerably more differentiated. But it is sufficient to explain the basic relations of authority (as compared to force) that is important for understanding rules.

<sup>10</sup> Wolff, *In Defense of Anarchism*, 5.

suggest that for Wolff, it is precisely the earlier-mentioned features of how commands are meant to be taken – exclusion and content-independence – that make obedience impermissible. According to Wolff, treating a command as a content-independent, exclusionary reason would violate our autonomy as individuals. If we had an obligation to do whatever another person demanded of us, this would prevent us from acting upon our own judgment of whether the demanded action is morally right or wrong. Since our highest moral duty is to follow our own independent judgment of whether an action is morally right or wrong, Wolff argues that no authority is ever legitimate – including the state itself.

## 2 Liberal Theories of Political Authority

One natural way of responding to Wolff's argument is to simply deny that autonomy is as important as he claims that it is. If we don't have an obligation to make up our own minds about important moral matters, or if that obligation is easily outweighed by other factors, then Wolff's argument has no force. And one might argue that the state certainly performs functions of sufficient importance to credibly purport to outweigh individual autonomy. It provides important social services, security, and so on, and as an empirical matter, the state might require the full cooperation of its members to carry out these functions.

However, this response is not available to the liberal tradition of political philosophy, the predominant tradition in contemporary Western political philosophy. Liberals as varied as Mill, Dworkin, Rawls and Raz all claim that autonomy is a central value, one that the state must respect at almost all costs. Liberals cannot deny that autonomy is important without leaving behind liberalism itself. Liberals must find a way of *reconciling* authority and autonomy if they are to recognize such a thing as legitimate authority. In this thesis, I argue that none of the attempts at reconciliation found in the leading liberal theories of authority are successful.

The first and most straightforward attempt that I will discuss (and reject) invokes what I will call *moralism* about political authority.<sup>11</sup> Many liberals see liberalism as derived from a set

---

<sup>11</sup> I borrow the term "moralism" from Bernard Williams, but differ in usage from him. For Williams, the point of introducing "moralism" is to draw a contrast with "realism" in political philosophy. Rejecting views that tie political

of moral values, like fairness, justice or even autonomy itself. Some of these liberals use these moral values to argue that citizens have political obligations, like a *prima facie* duty to comply with the law.<sup>12</sup> Moralism about political authority uses the same strategy to argue for the more demanding conclusion that citizens have a moral requirement to treat the law as a content-independent, exclusionary reason. That is, moralism uses (liberal) moral values to show that the law has legitimate authority. Few philosophers seem to have defended this view despite its apparent simplicity. It is important to see why it fails in order to understand why the liberal faces such a difficult task in accounting for legitimate political authority.

I begin by considering an advantage of moralism: if moralism was correct, and there was a moral requirement to treat the law as a source of content-independent, exclusionary reasons, we would be able to reconcile authority and autonomy. As we saw, Wolff thinks that our duty to be autonomous is a duty to act on our moral duties, as we understand them. But if an agent can plausibly reason that their moral duty *is* to obey the state, because such obedience is required by fair play or some other liberal value, then there would be no tension between legitimate authority and autonomy after all. That is, if the way that an agent fulfills their Wolffian duty of autonomy just is by fulfilling their moral duties as they see them, and if one of these moral duties is that they must treat laws as content-independent, exclusionary reasons, then they will have fulfilled their duty of autonomy while simultaneously submitting to the state's authority. The moralist would say that Wolff was correct about the value of autonomy, and correct that we must act on our assessment of what we think is morally correct, but that he failed to account for the possibility that that assessment might be to exclude some of our reasons.

However, while moralism provides an adequate explanation for why citizens have a moral requirement to treat laws as content-independent reasons, the explanation breaks down when we consider why they should treat laws as exclusionary reasons. As we have seen, this means that they should not act on certain reasons for action. This does not necessarily mean that

---

theory too closely to moral theory, Williams suggests that liberal political theory should instead “shape its account of itself more realistically to what is platitudinously politics”. (Williams, “Realism and Moralism” in *In The Beginning was the Deed*, 12)

<sup>12</sup> A *prima facie* duty to obey the law is a duty to obey the law, at least when it does not contradict other moral duties.

the law must exclude *every* other reason for action. Even in the drill sergeant example, soldiers remain free to act on certain reasons for action, like reasons relating to the avoidance of war crimes, and presumably the law leaves its subjects with an even wider pool of non-excludable reasons. Yet it seems clear that the state's authority presumably requires them not to act on *some* of their reasons, including some of their moral reasons. Despite the somewhat limited scope of this exclusion, I argue that moralism still fails to explain why an obedient person should not act on moral considerations that compete with what the command tells them to do. At most, moralism shows that citizens are required to treat commands or laws as ordinary moral reasons that sometimes outweigh competing moral commitments. But moralism does not show how laws or commands make those competing moral commitments cease to matter in the way that is characteristic of exclusion about how to act.

Moralism, at least in the simple form presented in this chapter, is a useful foil for seeing why liberal philosophers needed to go beyond it in their explanations of how legitimate political authority is possible. As we will see, with justice as fairness, John Rawls can be read as trying to sidestep the problems of simple moralism by introducing a two-step argument: liberal moral commitments are given by the original position, and then the original position is used to determine exclusionary principles of political authority. With the service conception of authority, Joseph Raz tries to skirt the problems of moralism altogether by arguing that authority performs the service of helping subjects to act on their reasons, moral and non-moral alike. And finally, with the second-person standpoint, Stephen Darwall seeks another ground that at least initially appears distinct from morality: the intrinsic force of our relationships to one another.<sup>13</sup> However, I will argue that none of these more complex theories are successful. This leaves liberals without an adequate ground for legitimate political authority. Some liberals may welcome this conclusion with relief, but I suggest that it is not without cost. It leaves liberal states without an important means of achieving coordination. Moreover, it leaves liberalism with an exposed flank. Those who are committed to legitimate state authority may treat the argument of this thesis as a *modus*

---

<sup>13</sup> It is somewhat less clear that Darwall is a liberal philosopher, since he does not tie his work on authority to any particular vision of the state. However, given his contractualist leanings, we can say his work is liberal insofar as the social contract tradition is liberal.

*tollens*: because liberals cannot account for the state's legitimate authority, liberalism must be mistaken.

### 3 Survey of the Thesis

#### *Chapter 2: Wolff's Challenge to Authority*

In this chapter, I present Wolff's challenge to legitimate authority. I begin by presenting his view of autonomy, focusing on the duties regarding deliberation and action that he thinks apply to any autonomous agent. I then discuss his view of authority, focusing on the features of authority and obedience that are relevant to its clash with autonomy. The primary problem with obedience, for Wolff, is that it requires us to not act on certain factors independent of the command itself that make a commanded action right or wrong. For example, if a citizen judges that a law is unfair, legitimate authority would require them to not act on that consideration at all. The secondary problem with obedience is that it requires us to treat content-independent considerations as reasons to act (in this case, commands) even though those considerations do not appear to be reasons at all. He furthermore argues that consent to an authority would be impermissible because it would simply amount to a willful forfeiture of autonomy.

#### *Chapter 3: Moralism*

In this chapter, I introduce the most direct way that the liberal might attempt to reconcile their commitment to autonomy with political authority. This is what I will call *moralism* about political authority.

I begin by considering liberal arguments for the existence of a *prima facie* duty to obey the law, a duty to obey the law as long as it does not conflict with other pressing moral values. These views attempt to derive the existence of a *prima facie* duty to obey the law from liberal moral values like fairness or even autonomy itself. I then ask whether the same strategy might be used to establish the existence of legitimate political authority.

At first glance, liberal view of political philosophy would offer an effective way of addressing Wolff's challenge. As we have seen, Wolff thinks that our duty to be autonomous is a duty to act on our moral duties, as we understand them. But if an agent must reason that their moral duty *is* to obey the state, then there is no tension between legitimate authority and autonomy after all. The moralist obeys the state *because of* their moral deliberations, not despite them. As long as they are grounded in moralism, then, liberal theories of authority appear to be safe from Wolff's challenge.

I argue that moralists have a satisfactory explanation for why citizens should treat laws as content-independent reasons. However, their explanation breaks down when trying to show that citizens should treat laws as exclusionary reasons. As we have seen, in the definition of legitimate authority adopted by Wolff, obedience involves not treating some factors that the agent has deemed to be morally relevant as reasons for action. However, moralist views of authority do not show why an agent should not treat these factors as reasons for action. To see what I mean, consider a moralist view that attempts to ground legitimate political authority on fairness. This would mean that fairness tells me to obey the law insofar as others in my society are doing the same. But now suppose that the state was distributing benefits and burdens on a society in an unfair manner. This seems to imply that fairness itself is pointing me in two opposing directions: I should obey the law (out of fairness), but I should disobey the law (also out of fairness). It is hard to see how considerations of fairness are going to give me permission to exclude the latter consideration from my deliberations. If my commitment to obey the law is derived from fairness, it seems that I cannot coherently disregard the value of fairness in the same context. Such incoherence would mean that I would not be fulfilling my duties as an autonomous agent to deliberate in an appropriate way.

#### *Chapter 4: Justice as Fairness*

In this chapter, I argue that John Rawls's "justice as fairness" does not satisfactorily address Wolff's problem. In short, justice as fairness is a system of liberal political morality that Rawls progressively develops in his works *A Theory of Justice* and *Political Liberalism*. It is a set of principles for a society and its members that is derived from the original position, a thought experiment in which agents choose principles for themselves and their institutions while lacking certain information about the position they will occupy in that society.

I begin by presenting the theory of political authority and political obligation found in *A Theory of Justice*, which involves a complex, multi-stage application of the original position. As a social contract theory, it might initially appear to be a form of consent theory, and if so, it would be unacceptable to Wolff for that reason. However, Rawls uses the idea of a social contract as a device for revealing the “fair terms of cooperation” and does not rely on the binding force of consent to explain the authority of the revealed principles. Instead, I think we should understand justice as fairness as a very intricate moralist theory, one that has two initial advantages over simpler moralist theories. First, justice as fairness can be interpreted in Kantian terms, and thus Rawls clearly expects that his theory of political authority is consistent with Kantian autonomy. Surely if any moralist theory can reconcile authority and autonomy, it would be a Kantian one. Second, Rawls has a novel, two-step strategy for explaining why citizens must exclude their own moral judgments when obeying the law. At step one, liberal moral commitments are given by the original position. At step two, the choices made by agents in the original position determine the principles that will govern their society’s institutions and individuals, including principles of political authority. Rawls argues that people in the original position would choose to be bound by principles of political obligation that require authoritative exclusion as a way of ensuring the stability of the principles of justice for their society’s institutions.

Despite these two advantages over simpler forms of moralism, I argue that justice as fairness does not successfully reconcile authority and autonomy. While the original position is a plausible way of construing autonomous reasoning, I argue that Rawls is incorrect that people in the original position would choose to be bound by a principle that requires them to not act on certain moral considerations that they deem relevant. In short, they would choose to leave it open to themselves to judge whether the loss of stability from disobeying an unjust law outweighs the injustice of the law, rather than excluding any of these considerations.

In *Political Liberalism*, Rawls instead presents justice as fairness in a different light. Rather than offering it as a correct moral view, or as a competitor to other moral views, he now proposes it as a basis for an “overlapping consensus” between reasonable moral views (or “comprehensive doctrines”). Rawls now hopes to show that justice as fairness, presumably including its principles of political authority, is acceptable to each reasonable comprehensive doctrine in circulation within a society. I argue that this new presentation of justice of fairness

simply pushes the problem of explaining authoritative exclusion onto the individual comprehensive doctrines. That is, we still need an explanation of how any comprehensive doctrine (e.g. justice as fairness) could require its holder to not act on some moral considerations.

#### *Chapter 5: The Service Conception of Authority*

In this chapter, I present Joseph Raz's service conception of authority. For Raz, authorities provide a service for their subjects: helping them act on their other reasons. For example, Raz would say that insofar as drivers have reason to get where they are going safely and quickly, they must accept the authority of a traffic light. Altogether, authority is ultimately a tool for helping us conform to right reason, and the pressure to obey it comes from the pressure to conform to right reason in our actions. Even if there were no moral obligations at all, the service conception would say that there still could be legitimate authorities.

The great achievement of the service conception of authority is that, unlike moralism, it makes sense of why someone should treat a command, a mere expression of someone's will, as an exclusionary, content-independent reason. In short, if such exclusion helps the subject conform to right reason in general, they would be irrational to act on the excluded reasons, and they should obey regardless of the content of what is commanded, because otherwise they will not get the benefits of obedience. However, as I argue in the next chapter, this ultimately is not enough to solve the problem of authority. It shows that obedience involves something more than treating commands as exclusionary, content-independent reasons.<sup>14</sup>

#### *Chapter 6: Darwall's Criticism of the Service Conception*

In this chapter, I present Stephen Darwall's objection to the service conception of authority. In short, Darwall argues that while rationality gives rise to its own kind of exclusion, the service conception fails to capture the *kind* of exclusion that is distinctive of practical authority in

---

<sup>14</sup> To clarify, this does not imply that Hart is mistaken when he identifies content-independence and peremptoriness (or exclusion) as features of how commands are meant to be taken. This is because he does not claim that they are *constitutive* of how they are meant to be taken – there might be something else that a subject needs to do to respond to commands as they are intended. Nor is this a problem for Wolff, who says that obedience involves doing what an authority says because they say so. Wolff does not claim that these are the constitutive features of obedience, although he might be taken to say that they are the only *morally impermissible* features of obedience.

general. For Darwall, legitimate authority makes it truly “illegitimate” to act on some of the features of a situation that seem morally relevant. Even if it would be irrational for me to disobey an authoritative command, as would be the case when the conditions of the normal justification thesis are satisfied, this does not render my disobedience illegitimate.

There are two main examples that illustrate Darwall’s objection. The first example is an expertise case, where the normal justification thesis appears to be satisfied because the subject lacks the expertise that would enable them to conform to right reason on their own. The second example is a motivation case, where the normal justification thesis is satisfied because the subject lacks the motivation to do as reason demands them to do. Darwall argues that neither case involves authority, and that the latter does not even involve a power to give people new reasons. In his replies to Darwall, Raz treats these two examples differently. Raz argues that the expertise case does not satisfy the normal justification thesis after all, and so it is no longer a counterexample to the normal justification thesis. With the motivation case, Raz doubles down on his claim that there really is exclusionary reason to obey the hired helper. I argue that Raz’s replies are not persuasive.

Finally, I introduce Darwall’s view of what *would* be required in order to make it illegitimate to contemplate the reasons: some sense in which an authority has a “right” to rule. I say more about his view of authority in the next chapter, but in this chapter I argue that Raz’s attempt to argue that the normal justification thesis does give rise to a right to rule is unsuccessful.

Overall, the most important lesson to draw from Darwall’s criticism of the service conception is that Raz’s attempt to ground legitimate authority in structural principles of practical reason is not successful. While Raz is not wrong to think that the principles of practical reason give rise to their own kind of exclusion, this is the kind of exclusion that is involved in advice rather than the kind that is involved in authority. As Hobbes might put the point, Raz confuses counsel and command. Despite explaining why subjects might treat commands as exclusionary, content-independent reasons, the service conception of authority is thus neither an account of *de facto* authority nor legitimate authority. The bindingness of authority is found in what is being reasoned about, not how we reason about it.

In this chapter, I consider Darwall's own "second-personal" view of authority. His main insight might be put in this way: without the will of other people to whom we are accountable, relatively little, if anything at all, is truly binding on us. He might say, for example, that if someone were the only person in the world, they would have no binding obligations at all, that they would be completely free to do as they like. This insight has important implications for authority. He says that for authority to actually bind its subjects, for it to actually exclude their competing reasons, those subjects must be accountable to the authority. That is what it means for an authority to have the "right" to make demands of its subjects.

One of the main reasons why Darwall's view of authority is so interesting is that it seems at first glance to offer an alternative to grounding political authority in morality or practical reason. We have already seen in the previous chapter that Darwall rejects grounding authority purely in terms of structural principles of practical reason. There are also good reasons to hope that second-personal authority is separate from morality. Not all interpersonal demands or relationships of accountability are matters of morality, yet they do seem to bind those to whom they apply. However, despite this initial hope, I argue that second-personal authority does turn out to collapse into either a form of moralism or a form of consent theory (which would also be unacceptable to Wolff).

I begin by explaining how Darwall's view of authority is part of a broader effort to shed light on "second-personal" aspects of our lives – cases where we address one another in a way that presupposes a relationship of accountability and respect, including any case in which we have a right to make a demand of another person. As Darwall understands authority, it is a relationship of accountability between two people in which the will of the person or body issuing the demands, or at least whomever it represents, play an ineliminable role in the creation of the obligation to do as it says.

Darwall says relatively little about the asymmetrical relations of authority like those between a state and its citizens, and much of my effort in this chapter is directed towards reconstructing what he would say about these relations. I suggest that there are two ways of extending the second-personal view of authority into a theory of political authority.

The first theory simply applies the general second-personal definition of authority to the particular case of political authority. Just as one person can be accountable to another person, the

first theory would say that a citizen is accountable to their state. However, I argue that this theory collapses into moralism or consent theory. In other work, Darwall also describes legal authority as a power to “find” or “make” law. This suggests a second theory of authority, one that implies that political authority involves not just a right to make claims and demands that are owed to the person who makes them, but also a power to *create* bipolar obligations that are owed to third parties, as well as moral obligations more generally. However, Darwall says little about how such authority would be justified, and what he does say on this subject appears to also lead him back to moralism.

Altogether, then, the second-person standpoint does not turn out to be a genuinely different justification for why autonomous agents should obey legitimate authorities. It collapses into moralism or consent theory, and otherwise seems unsuitable for a liberal society.

### *Conclusion*

After summarizing the argument of the thesis, I close with two observations. First, it may not necessarily be so undesirable that liberal theory lacks an adequate account of political authority. Perhaps a liberal society can get by without its members having a duty to obey the law just because it is the law. However, I also note some disadvantages of this conclusion. Second, I note that my argument is not exhaustive. I have considered many of the major liberal theories and argued that none of them provide a standpoint that can make sense of why someone might be permitted to not act on moral considerations that seem relevant to them, but this does not mean that there is no such standpoint. I close by suggesting that arbitration might provide one such standpoint, drawing on some work by Daniel Viehoff, who argues that arbitration provides a service to its subjects, but one different than the service suggested by Raz’s original service conception.

## Chapter 2 Wolff's Challenge to Authority

### 1 Introduction

In 1970, Robert Paul Wolff published *In Defense of Anarchism*, in which he drew the dramatic conclusion that all states are fundamentally illegitimate. Up until this point, the arguments that the state lacked legitimate authority tended to take one of two forms. Some arguments drew upon cases which generate clear intuitions that the state has no legitimate authority over us, like the famous case of the well-lit stop sign in the middle of the night. Other arguments methodically showed the failure of common arguments once thought to establish the state's legitimate authority (like fair play or gratitude).<sup>15</sup>

Wolff intended to offer a more fundamental and far-reaching argument against the legitimate state. According to Wolff, all states are illegitimate because their claim to legitimate authority clashes with our autonomy as individuals. The state claims that we have an obligation to do whatever it demands of us, no matter what it demands of us, in a way that excludes our own judgment of whether the demanded action is morally right or wrong, and that requires us to substitute its judgment instead, at least regarding how we ultimately act. Since our highest moral duty is to follow our own judgment of whether an action is morally right or wrong, Wolff argues that no state is ever legitimate, and that none of its exercises of authority are legitimate or binding.

One natural way of responding to Wolff's argument is to simply deny that autonomy is as important as he claims that it is. If we don't have an obligation to live according to our own judgments about important moral matters, or if that obligation is easily outweighed by other factors, then Wolff's argument has no force. However, this response is not available to the liberal tradition of political philosophy, the predominant tradition in contemporary Western political

---

<sup>15</sup> Many of these arguments were framed in terms of the duty to obey the law rather than legitimate authority, but as we will see, Wolff understands the duty to obey the law and legitimate authority to be closely related to one another.

philosophy. Liberals as varied as Mill, Dworkin, Rawls and Raz all claim that autonomy is a central value, one that the state must respect at almost all costs. Liberals cannot deny that autonomy is important without leaving behind liberalism itself.

Wolff's argument thus poses a deep challenge for liberal theory. Liberals must find a way of *reconciling* authority and autonomy, if they are to recognize such a thing as legitimate authority. This challenge forms the basis for the rest of the thesis. As we will see in subsequent chapters, the leading liberal attempts to meet this challenge have not been successful.

## 2 Wolff's View of Autonomy

I will begin my explanation of Wolff's argument by discussing his view of autonomy. In his initial explanation of autonomy, he borrows heavily from the Kantian concept of autonomy. He approvingly summarizes Kant's view of autonomy as one that treats autonomy as a "combination of freedom and responsibility".<sup>16</sup> He seems to imagine autonomy as a state in which we have achieved the maximum potential of our agency. While as agents, we are always responsible for our actions, being an autonomous agent involves the further step of *taking* responsibility for our actions. When he first discusses autonomy, he suggests that this taking responsibility has a negative component ('not being subject to the will of another') and a positive component ('submission to laws which one has made for oneself'), both of which reflect concepts and terms found in Kant's moral philosophy.<sup>17</sup>

However, Wolff soon leaves behind Kantian theory.<sup>18</sup> For Kant, moral autonomy is important as a precondition to moral agency itself – it is a presupposition of deliberation, rather than something that we deliberate about. Wolff instead frames the duties relating to our autonomy in more broadly appealing terms of deliberation, decision-making and action.

---

<sup>16</sup> Wolff, *In Defense of Anarchism*, 14.

<sup>17</sup> Wolff, *In Defense of Anarchism*, 14.

<sup>18</sup> For a detailed argument that Wolff's view is not Kantian, see Riley, "On the "Kantian" Foundations of Robert Paul Wolff's Anarchism".

First of all, Wolff says that we have to deliberate about what to do – that is, we have to form a reasoned judgment about what to do. Wolff suggests that this judgment must take into account “all of the available information” that concerns the possible actions we might perform in a situation.<sup>19</sup> Which information are agents required to take into account when in forming that judgment? Part of the answer can be read into the fact that for Wolff, a central part of deliberation seems to involve the application of moral principles to particular situations. Thus, the main duty regarding deliberation is to form a reasoned view about which moral principles are correct – that is, to “attempt to ascertain what is right.”<sup>20</sup> There is a further duty, one that involves learning the non-moral information that is required in order to apply moral principles to concrete situations. Wolff even goes so far as to say that American citizens are obliged to “master enough modern science to follow debates about nuclear policy”, though he implies that such technical information is only relevant insofar as it is relevant to the real decisions that lie in front of us, like for whom to vote in an election.<sup>21</sup> The duty is weaker than the main one because Wolff believes that it is far more important to deliberate for ourselves about moral principles than technical matters. On the one hand, on certain technical questions, like technical medical questions that can only be answered by a doctor, Wolff suggests that it is “reasonable” to “give up his independence of judgment”.<sup>22</sup> But on the other hand, with respect to moral principles, Wolff insists that we alone must be “the judge of those constraints”.<sup>23</sup> The only respect in which we can rely on others for our knowledge of moral principles is by allowing them to point them out to us, so that we may verify them for ourselves, in much the same way that a mathematical proof can be demonstrated to an audience:

He may listen to the advice of others, but he makes it his own by determining for himself whether it is good advice. He may learn from others about his moral obligations, but only

---

<sup>19</sup> Wolff, *In Defense of Anarchism*, 46.

<sup>20</sup> Wolff, *In Defense of Anarchism*, 13.

<sup>21</sup> Wolff, *In Defense of Anarchism*, 17.

<sup>22</sup> Wolff, *In Defense of Anarchism*, 15.

<sup>23</sup> Wolff, *In Defense of Anarchism*, 13.

in the sense that a mathematician learns from other mathematicians – namely by hearing them from arguments whose validity he recognizes even though he did not think of them himself. He does not learn in the sense that one learns from an explorer, by accepting as true his accounts of things one cannot see for oneself.

What Wolff is saying in this passage is that people might be initially unaware of moral duties that apply to them, and another person can help to make them aware of those duties, but the individual remains the final judge over whether they do, in fact, have those moral duties.

Wolff also seems to think that an autonomous person must, when forming their judgment, neglect certain kinds of considerations that are genuinely irrelevant to decision-making. In particular, Wolff suggests that allowing chance (e.g. flipping a coin) to play a role in a decision would generally be a “willful forfeiture of autonomy”.<sup>24</sup> For example, if someone were to whimsically make important life decisions by flipping a coin, Wolff would regard that person as having violated their duty of autonomy.

Finally, Wolff notes that it is not necessarily the case that all information relevant to a decision is going to be available to us. While we do have an obligation to incorporate into our decision-making the information that is “theoretically within our control”, we do not have an obligation to incorporate information that is outside of our control, such as certain forms of inescapable “ignorance”.<sup>25</sup> Our obligation, then, is to integrate into our deliberation all relevant information that we do not face an insurmountable barrier to obtaining or using.

To summarize Wolff’s position so far, he says that we have a duty to form a considered judgment about what we should do in each situation – one that recognizes moral principles that we accept for ourselves, and one that excludes irrelevant information. But Wolff would not say that this is the full extent of the duty of autonomy. Autonomy requires us to do more than simply form a judgment about what to do. In order for us to be autonomous, this judgment must be manifested in our intentions and actions. First, the autonomous agent must be “committed” to the

---

<sup>24</sup> Wolff, *In Defense of Anarchism*, 47.

<sup>25</sup> Wolff, *In Defense of Anarchism*, 47.

judgment one has arrived at through deliberation.<sup>26</sup> Wolff seems to suggest that this involves willing an “imperative” that one expresses to oneself.”<sup>27</sup> Second, the agent must act only on the basis of such imperatives. Only when one acts on the basis of such a judgment does one count as having made the “final decision” about what one should do.<sup>28</sup>

With that summary in mind, one last point about Wolff’s view of autonomy is important to emphasize. While many other philosophers defend autonomy on instrumental grounds, like how it contributes to our well-being, or how it promotes other goods, for Wolff these duties to deliberate and act on the results of one’s deliberations are not instrumental. While it is probably true that someone who deliberates and acts on the results of their deliberation is more likely to act morally correctly, for Wolff, autonomy is not merely a means to morally correct behaviour. It has inherent value, such that even if someone were awful at moral reasoning, Wolff would say that they would still have overriding reason to continue to reason for themselves anyways.

### 3 Wolff’s View of Authority

The next step of Wolff’s argument is an analysis of authority, which he undertakes by looking at the claims made by institutions commonly thought to possess it, particularly states and legal systems. He notes that one of the essential features of any state or legal system is that it makes a distinctive claim of authority on those subject to it – it claims the right to command them.<sup>29</sup> Wolff begins his discussion of authority by observing that the authoritative claims made by states on their citizens are not those of “power”, understood as “ability to compel compliance,

---

<sup>26</sup> Wolff, *In Defense of Anarchism*, 39.

<sup>27</sup> Wolff, *In Defense of Anarchism*, 13.

<sup>28</sup> Wolff, *In Defense of Anarchism*, 15.

<sup>29</sup> Some philosophers, like Joseph Raz, use the description “*de facto* authority” to describe someone who makes this claim. Wolff, on the other hand, suggests that *de facto* authorities are authorities that are believed to have the right to command.

either through the use or the threat of force”.<sup>30</sup> Instead, legitimate authority is a matter of “right”. Nor are they like the claims of advice or theoretical authority. In the case of law, the law’s claim to authority implies that those subject to it have the “duty to follow the law just because the law says so”.<sup>31</sup> To recognize this claim to authority, and treat the authority as legitimate, is to obey the law just because the law says so.<sup>32</sup>

The key to understanding Wolff’s view of authority, then, is to make sense of what is meant by a “duty to obey the law because the law says so”. Wolff makes several remarks about this, which I will collect here. First, when we say that the state has legitimate authority, we do not mean only that there are already good independent reasons to do as the law says.<sup>33</sup> Thus, if the state passes laws that happen to be independently just, and the only reason that the citizen follows the laws is that they are independently just, then the citizen does not acknowledge that the state has legitimate authority. Second, Wolff writes that authority “resides in persons”, and that it is possessed “in virtue of who they are and not by virtue of what they command”.<sup>34</sup> Third, Wolff draws a distinction between an “authoritative command” and a “persuasive argument”. A persuasive argument is intended to persuade a person to do something by making them “believe that it is something that [they] ought to do”, and serves merely as the “occasion for becoming aware of [their] duty”, while neither of these things are true with respect to authoritative commands.<sup>35</sup>

Wolff’s example of a passenger of a sinking ship illustrates some of these points and further clarifies his view of obedience. In this example, the passenger of a sinking ship is confronted by the sight of the ship’s captain giving orders that direct passengers to the lifeboats.

---

<sup>30</sup> Wolff, *In Defense of Anarchism*, 4.

<sup>31</sup> Wolff, *In Defense of Anarchism*, 9.

<sup>32</sup> While Wolff primarily discusses the state’s claim to authority, note that he would probably grant that there are others who make claims of authority – like parents or other figures.

<sup>33</sup> Wolff, *In Defense of Anarchism*, 9.

<sup>34</sup> Wolff, *In Defence of Anarchism*, 6.

<sup>35</sup> Wolff, *In Defence of Anarchism*, 6.

Wolff then describes a way of taking those orders into consideration that he explicitly says does *not* amount to obedience.<sup>36</sup> The passenger might weigh many different factors, including the obvious factors like the disorder caused by deviating from the orders, but also the less obvious factors, like whether he or another person is more deserving of the lifeboat. According to Wolff, taking the captain's orders into consideration in this way doesn't amount to obedience, because "insofar as [the passenger] makes such a decision, [he] is not obeying his command; that is, [he] is not acknowledging him as having authority over [him]."<sup>37</sup> To turn the situation into one that actually involves obedience, it would be necessary for the passenger to *not* engage in this process, to not weigh the circumstantial factors of getting into the lifeboat or not getting into the lifeboat in deciding what to do.

Altogether, it seems that obedience to authority requires *someone to disregard their judgment of what they think is morally right, and instead to do what the authority demands, regardless of what it is.*<sup>38</sup> For example, in the lifeboat case, the obedient person disregards their judgment that deviating from the captain's orders might cause disorder and that it might be more important to save children than adults, and simply obeys the captain's orders. They treat the captain's order as a command, not merely as advice, and their reason for doing so is because he is the captain, not because of what he ordered. Authority, then, is a right to this kind of obedience.

Because this explanation of authority and obedience is so crucial to the argument of the thesis, and because the argument of the thesis might seem wrong-headed if the definition of authority that it adopts is widely off the mark, I will supplement Wolff's argument and try to show that he is correct that the state's claim to legitimate authority should be understood as a demand that its members act because it says so.

---

<sup>36</sup> Wolff, *In Defence of Anarchism*, 16.

<sup>37</sup> Wolff, *In Defence of Anarchism*, 16.

<sup>38</sup> This foreshadows my discussion of Raz later in the thesis. As we will see, the idea of an exclusionary reasons for action is central to Raz's view of authority. At this point, however, I am not committed to saying that exclusion involves a *reason* to exclude, merely that someone who obeys an authority *does* exclude these considerations from their deliberation.

First, it seems that Wolff is correct that the intuitive force of obedience involves some kind of disregard for one's own judgment of the situation, or at least a disregard of aspects of that judgment. To see why, consider the classic case of the traffic light. Intuitively, a traffic light doesn't merely give a driver strong reason to do as the traffic light demands. Instead, it seems like the traffic light is intended to exclude considerations that speak against what it requires. When the traffic light is red, the driver isn't supposed to think about how quickly one would like to get to one's destination, how many pedestrians are in the intersection, or even whether there is oncoming traffic. The driver is simply supposed to stop. (That said, the traffic light does not exclude *every* consideration. If it is an emergency, or the traffic light is broken, perhaps I become free to disregard its instructions. But that doesn't seem to stand in the way of the traffic light counting as authoritative in some sense.) The state, for Wolff, is a traffic light *writ large*.

I do not think that there are any suitable alternatives to exclusion and content-independence for explaining this intuitive force. One might think that authorities simply 'drown out' a subject's reasons by giving them a reason that is sufficiently large as to overwhelm their other, pre-existing reasons. However, I doubt that this is a fitting description of authority, because such a picture of practical reasoning would be considerably coarser than the picture provided by exclusion. For example, commands can exclude some reasons for action, but not all, but the 'drowning out' picture does not seem to allow for this possibility. Another alternative is that authorities affect the practical reasoning of those subject to them by altering their other reasons. Larry Alexander suggests that law "affects the balance of first-order reasons by affecting how considerations of coordination, disobedience, and sanctions play out on that balance."<sup>39</sup> For example, one might say that the traffic light serves to trigger a coordination scheme in which all parties have the mutual knowledge that all other parties are aware of this coordination scheme. However, this explanation seems to proceed in exactly the wrong direction: it complicates, rather than simplifies, the practical reasoning that is required to approach a traffic light. Finally, one might think that the traffic light is an exception, and that the law in general does not have this force. But this would require a likely ad hoc explanation of why the traffic

---

<sup>39</sup> Alexander, "Law and Exclusionary Reasons," 8

light is special, and how subjects could successfully distinguish between the cases where the law has exclusionary authority and cases where it has some other kind of authority.

Second, I acknowledge that many philosophers use the term “legitimacy”, and even “legitimate authority”, in a different way than Wolff. Some use “legitimacy” in a way that has nothing directly to do with what a subject has obligation to do, for example, to describe the circumstances in which the state is justified in using force. Other philosophers have substantive, rather than content-independent, theories of legitimacy. They would say that a state is legitimate only if it fulfills certain tasks or respects certain fundamental principles, and that the fact of the state’s legitimacy entails nothing about what subjects must do. In contrast, Wolff is saying that legitimate political authority is correlative to political obligation – if there is a duty to obey the law just because it is the law, then the state has legitimate authority, and vice-versa. To some degree, these differing views of legitimate authority might simply reflect a semantic disagreement – there might be more than one characterization of “legitimacy”, each an answer to a different question in political philosophy. However, to the extent that Wolff seeks to isolate a single important concept of legitimacy, one associated with the duty to obey the law, note that he is not alone in treating obligation and legitimate authority as correlative – many other philosophers, including Joseph Raz, agree that there is a close relationship between legitimate political authority and political obligation. Moreover, if it turns out that this view of authority is too demanding to actually be morally justifiable, and that no state is actually legitimate, Wolff would presumably say that this is not a problem for his view, but instead simply indicates that many of our initial beliefs about state legitimacy were mistaken.

Third, note that it is not essential to Wolff’s argument that *every* law carries this obligating force. Wolff could consistently say that in a legitimate state there is a duty to obey certain laws, like laws about matters of pressing public interest, but not others, like ordinary traffic laws. What matters is that the state has the *right* to have its laws taken in this way, even if it does not exert this power on every occasion that a law is passed.<sup>40</sup> Thus Wolff’s view of

---

<sup>40</sup> To defend this point in full, one would need an account of why the state obligates through law in some cases but not in others. One possibility is that the choice to enforce a legal command criminally rather than civilly signals the presence of this intention to obligate.

authority does not necessarily lead to the implausible conclusion that someone is acting immorally when they jaywalk, or that the possibility of moral jaywalking undermines the state's legitimate authority. Moreover, as we will see, Wolff's argument against authority still applies to limited claims of authority, like authority over a relatively narrow range of issues.

## 4 The Primary Conflict Between Authority and Autonomy: Exclusion

Wolff believes that obedience will inevitably violate the duties to deliberate and act on the basis of that deliberation. He provides his primary statement of the conflict between authority and autonomy in the following passage:

The defining mark of the state is authority, the right to rule. The primary obligation of man is autonomy, the refusal to be ruled. It would seem, then, that there can be no resolution of the conflict between the autonomy of the individual and the putative authority of the state. Insofar as a man fulfills his obligation to make himself the author of his decisions, he will resist the state's claim to have authority over him. That is to say, he will deny that he has a duty to obey the laws of the state simply because they are the laws.<sup>41</sup>

Later in *In Defense of Anarchism* he suggests that this incompatibility leaves the agent with only two options:

Either we must embrace philosophical anarchism and treat all governments as non-legitimate bodies whose commands must be judged and evaluated in each instance before they are obeyed; or else, we must give up as quixotic the pursuit of autonomy in the political realm and submit ourselves (by an implicit promise) to whatever form of government appears most just and beneficent at the moment.<sup>42</sup>

---

<sup>41</sup> Wolff, *In Defense of Anarchism*, 18.

<sup>42</sup> Wolff, *In Defense of Anarchism*, 71.

Since it is “out of the question to give up the commitment to autonomy,”<sup>43</sup> Wolff thinks that we should take the first option and embrace philosophical anarchism.<sup>44</sup>

These two statements do not explicitly say *why* Wolff thinks that authority and autonomy are in conflict with one another, but the primary source of this conflict seems fairly clear. An obedient person does not act on some of the considerations they have raised in their deliberations. They treat things that they regard as reasons for action (the aspects of the situation they deemed to be morally relevant about what to do) as non-reasons. Following Raz, we can call this quality the *exclusionary* nature of legal commands, because they exclude at least some reasons from action.<sup>45</sup> We can illustrate exclusion with the following simple model of deliberation: ordinarily in deciding whether to do an action, an agent will take all of the reasons that speak for or against it, add them up, and then either do the action depending on which way the balance of reasons tilts. When a reason is excluded, it is simply not added to the final sum, and so plays no role in determining how the agent ultimately acts. Wolff is saying that it is a grave error for an agent to exclude reasons in this way, at least in determining how they ultimately act.

---

<sup>43</sup> Wolff, *In Defense of Anarchism*, 71.

<sup>44</sup> Wolff’s conclusion that there is a conflict between authority and autonomy has one self-admitted exception. This exception is that he suggests that there is (after all) a form of government that offers the prospect of reconciling authority and autonomy. This is unanimous direct democracy, a political community in which “every person votes on every issue – governed by a rule of unanimity” – that is, a community in which every person would have a veto over every proposed law. (Wolff, *In Defense of Anarchism*, 23) In a state governed by unanimous direct democracy, because every law passes with the support of each voter, it follows that “every member of the society wills freely every law which is actually passed”, which implies that the laws count as “dictates of his own will”. (Wolff, *In Defense of Anarchism*, 23) That said, it seems clear that unanimous direct democracy is not a genuine solution to the problem of political authority. If we really do have a duty to deliberate, it follows that each time a citizen is confronted with a law, they must “rehearse to themselves the arguments which persuaded him to vote as he did yesterday, and he will either find these still convincing or no longer convincing.” (Reiman, *In Defense of Political Philosophy*, 13) If the only government that is consistent with autonomy is direct democracy where individuals have a “continual” right of veto, that surely cannot count as authority at all. (Reiman, *In Defense of Political Philosophy*, 15) And in other work, Wolff ultimately concedes this point. He grants that the unanimous direct democracy is problematic, since it does not “create the sort of political authority that [he] was attempting to analyze.” (Wolff, “Reply to Reiman,” 3) Wolff appears to anticipate precisely this objection in *In Defense of Anarchism*, where he writes “It might be argued that even this limiting case is not genuine, since each man is obeying himself, and hence is not submitting to a legitimate authority.” (Wolff, *In Defense of Anarchism*, 23)

<sup>45</sup> Raz, *Practical Reason and Norms*, 184-5.

The conflict between autonomy and authoritative exclusion is made more severe by Wolff's earlier observations about the structure of deliberation. For Wolff, the finer structure of deliberation involves moral deliberation about moral principles, and then the application of those principles to the situation at hand using empirical and technical information. In many cases of obedience (though perhaps not all) an obedient person can be said to have violated the moral principles that they took themselves to endorse. If obedience requires someone to treat some feature of a situation as not relevant to action, but their moral principles say that it is such a feature, the obedient person will have violated their guiding moral principles.

## 5 The Secondary Conflict Between Authority and Autonomy: Content-Independence

There is also a secondary source of the conflict between authority and autonomy. Not only does obedience involve treating genuine reasons for action as non-reasons, it also involves treating a command (legal or otherwise) as a reason for action. Wolff writes the following:

I also deny that there are, or could be, states which are legitimate in the sense that the validity of their laws constitutes, in itself and independently of considerations of long-run consequences or side-effects, a reason of any weight at all for complying with those laws. .... But when all this has been weighed in the balance on one side or the other, if someone says to him, "Furthermore, this is a *legitimate* state (because it is a genuine democracy, or because it is the dictatorship of the proletariat, or because it is God's vicar on earth, or whatever), and that fact should count by itself for something in your deliberations"; then I say, no, that is never a good reason – that is never a reason which deserves to be given any weight at all.<sup>46</sup>

In this passage, Wolff denies that the legitimacy of a command ever "deserve[s] to be given any weight at all". Presumably, he thinks that it follows from this that someone who does treat a command as a reason is violating their duties of autonomy. This raises the question: why does it

---

<sup>46</sup> Wolff, "Reply to Prichard and Smith," 303.

violate the duty of autonomy to treat a command as a reason? Wolff's answer appears to be that treating someone's commands as reasons amounts to an illegitimate "double-counting" of reasons.<sup>47</sup> That is, he seems to think that the only reason to treat a command as a reason is because there is already some underlying moral factor, aside from legitimacy itself, that explains why the command should be treated as a reason, and to treat the fact that something is commanded as a new reason would fail to take into account that underlying factor. However, this "double counting" argument is ultimately circular. To say that treating commands as reasons is "double counting" is just to presuppose without argument that the fact that something is commanded does not matter, morally speaking. And either way, Wolff would need to show that this double counting is not merely mistaken, but that it is so mistaken as to be a violation of the duty of deliberation. Not every mistake in deliberation will violate the duty of autonomy, since Wolff does not suggest that an agent must deliberate correctly in order to fulfill their duty of autonomy.

Occasionally, Wolff suggests a different, more directly moral problem with treating commands as reasons to do what is commanded. According to this answer, it is not that someone would be mistaken, or deliberating badly, to treat another person's commands as reasons, but rather that doing so has grave moral consequences. For example, it might be intrinsically wrong for an agent to treat a person's commands as reasons for action because doing so would make them subject to that agent's will, even if doing so was rational and correct. Wolff hints at this in the following description of an autonomous person: "The autonomous man, insofar as he is autonomous, is not subject to the will of another. He may do what another tells him, but not because he has been told to do it. He is therefore, in the political sense of the word, free."<sup>48</sup> I agree with Wolff that this submission to another person's will might be a problem, but it is not clear to me that this is central to his main critique of authority. For example, Wolff does not worry about this in his analysis of the captain directing passengers to the lifeboat. While Wolff would say the passengers remain free to weigh the various advantages and disadvantages with obeying the captain's orders, it is easy to imagine how, by making the emergency particularly

---

<sup>47</sup> Wolff, "Reply to Reiman," 11.

<sup>48</sup> Wolff, *In Defense of Anarchism*, 14.

dire, those advantages and disadvantages will always tilt towards obeying him, at least within the limited range about what to do regarding the evacuation of the ship. In such a case, the passenger would be subject to the will of the captain; the captain would have the power to change what it is the passenger must do. Yet Wolff does not suggest that he finds such a power to be inherently objectionable. For Wolff, it appears that the essential problem with obedience has to do with its incompatibility with the proper form of deliberation and action. An obedient person is not living up to the deliberative demands of their agency.

We thus need a new explanation for why treating commands as reasons violates our duty of autonomy. To clarify this secondary conflict between authority and autonomy, I will draw on H.L.A. Hart's idea of *content-independence*. The content-independence of a command "lies in the fact that [she] may issue many different commands to the same or to different people and the actions commanded may have nothing in common, yet in the case of all of them the commander intends [her] expressions of intention to be taken as a reason for doing them."<sup>49</sup>

Commands are intended as both exclusionary and content-independent reasons, but to see how exclusion and content-independence might come apart, note that someone might have an exclusionary reason to not act on certain reasons, without any reason to replace them. This is arguably the case in Joseph Raz's example of an exhausted person making an investment.<sup>50</sup> Such a person has reason not to act on some of their reasons, but there is no reason that comes into play to replace them. Likewise, there are many examples of reasons that are content-independent (at least as defined by Hart) without being exclusionary. For example, on one interpretation, requests arguably take this form.<sup>51</sup> The fact that someone requests something is a content-independent reason to do it, more or less regardless of what is being requested, but does not exclude any other reasons. It is simply added to the other reasons that already speak for or against doing the action.

---

<sup>49</sup> Hart, *Essays on Bentham*, 254.

<sup>50</sup> Raz, *Practical Reason and Norms*, 37.

<sup>51</sup> Raz, *The Morality of Freedom*, 36.

By using the idea of content-independence, I think that Wolff could offer a somewhat more substantive explanation of why treating the mere fact that something is commanded as a reason violates the duty of autonomy. As we've seen, he thinks that there are certain ways of "deliberating" that are *so* catastrophic that they would fail to count as deliberating at all. He seems to be saying that one of these ways is treating a content-independent consideration as a reason to do some action. To act on a content-independent consideration, something that could have been a reason for doing any action, in the right circumstances, is a gross error in deliberation and action. The content-independent consideration bears no discernable relationship to the choiceworthiness of the action. Likewise, he would say that whether an action is commanded doesn't actually speak to the properties that make the action worth doing or worth avoiding.<sup>52</sup> This seems to also be why the earlier example of the coin flip involves a failure of deliberation. A coin flip has nothing to do with the properties that make the action worth doing or worth avoiding.

Let me briefly sum up Wolff's argument so far. Legitimate authority is a right to obedience – a right that someone does what another person demands, just because that person said so. Or in the case of the state, the state's legitimate authority would consist of its right that its citizens obey the law just because it is the law. There are at least two conflicts between authority and autonomy, both of which concern the way in which an obedient person is expected to deliberate and act. The primary conflict is that the obedient person excludes certain reasons from their deliberations about how to act. The secondary conflict is that an obedient person treats a command or demand – a content-independent consideration – as a reason to act.<sup>53</sup>

---

<sup>52</sup> One might object that the fact that some action is commanded *is* part of the content of the action. If so, it would follow that there is nothing problematic about treating a content-independent consideration, at least in Hart's sense, as a reason for action. This would imply that we need a more substantive account of content-independence, one that with a more metaphysically elaborate view of content, to capture what is problematic about acting on commands. For further discussion, see Sciaraffa, "On Content-Independent Reasons: It's Not in the Name".

<sup>53</sup> Scott Shapiro has suggested that there is nothing particularly wrong with acting on an exclusionary reason, or treating a command as a reason, in isolation from one another. Rather – and this is what is most intriguing about Shapiro's position – he thinks that the problem lies with how the two would function together in an agent's practical reasoning. (Shapiro, "Authority," 390)

One last point about Wolff's argument is important for appreciating why it is such a serious problem for political philosophy. Though Wolff notes that the state's claim to authority is supreme and universal, in that it claims to bind all citizens to all laws, his argument is a problem not only for supreme or universal claims to authority. It would also affect narrower claims to authority on particular matters. Even if the state only claimed legitimate authority over a single question – abortion, drug use, even traffic law – Wolff's argument would still apply to that claim. We would still be violating our duties of autonomy if we conceded to the state the right to determine for ourselves what to do on that case. (That said, given Wolff's explanation of the structure of the deliberation, and the role of moral principles in deliberation, it seems clear he would regard it as significantly worse to obey the state on purely moral questions like abortion or drug use than on matters that lack moral gravity.)

## 6 Promises and Decisions

Since the main claim of this thesis is that Wolff's challenge has not been addressed by liberal theory, I should first respond to an important objection that might prevent the challenge from getting off the ground in the first place. This objection concerns promises and decisions, and takes the form of a *reductio ad absurdum* argument: if Wolff is correct, and obedience is impermissible because it requires us to exclude certain considerations from our decisions about how to act, promises and decisions must also be impermissible for the same reason. They undermine autonomy in the same way as obedience. Since promises and (especially) decisions surely are permissible, the argument goes, Wolff's argument must be mistaken. As Leslie Green writes of promises, "Given the view that promises create new and independent obligations, it follows that any promise is a surrender of autonomy."<sup>54</sup> Likewise, Gerald Dworkin writes of decisions, "From the temporal perspective the commitments of my earlier self must bind (to some degree) my later self. It cannot always be open for the later self to renounce the commitments of the earlier self."<sup>55</sup>

---

<sup>54</sup> Green, *The Authority of the State*, 32.

<sup>55</sup> Dworkin, *The Theory and Practice of Autonomy*, 42.

In reply, I deny that those promises require an agent to exclude any moral reasons. I grant that promises may have a moderate exclusionary force – the making of a promise makes it the case that certain *prudential* factors no longer count in how one ultimately acts.<sup>56</sup> For example, if I have promised to pick up a friend, one might say that I am no longer permitted to take into account my desire to get home more quickly in deciding whether to follow through with that promise. But this is arguably true of any moral duty – moral duties in general have exclusionary force over prudential considerations. However, promises do not exclude any genuinely moral factors from how one must act; instead, they serve as additional moral reasons for action to be added to the situation. Promises serve as these reasons for action because of their relationship to other moral duties – for example, our moral duty to avoid harm and disappointment to others.

Decisions, in contrast, genuinely do require the agent to exclude considerations that arise between when the decision is taken and when action occurs, including moral considerations. (At least the making of the decision requires this. Once the decision is enacted, perhaps there is no longer an opportunity to reason at all.) However, without decisions there could be no action, and since part of the duty of autonomy is realizing the results of one’s deliberation in action, Wolff must accept this as a necessary evil.

## 7 The Liberal Response to Wolff

Though many philosophers have raised objections to Wolff’s argument, like the *reductio ad absurdum* objection mentioned in the previous section, the most common way of responding to Wolff’s argument is to simply deny that autonomy is as important as he claims that it is. For example, Gerald Dworkin argues that a “preoccupation with autonomy” is motivated purely by a mistaken “attribution of supremacy to the concepts of integrity and conscientiousness”.<sup>57</sup> If we don’t have an obligation to act in light of the conclusions that we draw about important moral matters, or if that obligation is easily outweighed by other factors, then Wolff’s argument has no

---

<sup>56</sup> See also Alexander, “Law and Exclusionary Reasons,” 15.

<sup>57</sup> Dworkin, *The Theory and Practice of Autonomy*, 40.

force. And the state certainly performs functions of sufficient importance to credibly threaten to outweigh individual autonomy – it provides important social services, security, and so on, and arguably requires the full cooperation of its subjects to do so.

However, this response is not readily available to the liberal tradition of political philosophy, the predominant tradition in contemporary Western political philosophy. As we will see, liberals as varied as Mill, Rawls and Raz all claim that autonomy is a central value, one that the state must respect at almost all costs, though different liberals have different reasons for emphasizing it. Some liberals will defend autonomy on the grounds that it is necessary for living a good life. Other liberals will defend autonomy on more deontological grounds, as essential to respecting ourselves and others as agents. In either case, however, liberals are committed to autonomy. One might object that Wolff's argument rests on a very particular, Kantian interpretation of autonomy that not all liberals will accept. However, Wolff is not particularly faithful to the Kantian origins of his argument, and it seems like there are good reasons to think that any liberal would value a life in which people deliberate fully about their actions, and act on the results of that deliberation, perhaps for much the same underlying reasons that liberals defend values like freedom of expression or conscientious objection.

Wolff also rejects another traditional liberal way of reconciling authority and autonomy: consent. Consent is traditionally invoked as a limit on state power, one that acts as a check on the state's ability to interfere with citizens, but the liberal might instead try to use consent to reconcile autonomy and authority. The state's power to demand that I do what it demands, just because it says so, might seem at first glance to be a terrible infringement on my autonomy, but what if I consent to being infringed upon in that way? Wouldn't that make it an *expression* of my autonomy, rather than a constraint on it, assuming that it was my free and rational choice to submit myself to the state's authority?

Wolff does not think so. He understands the force of consent in terms of its relation to promising. He grants that someone who promises to obey the state would find themselves under an obligation to obey the state, however he says that such a promise would occur "precisely by

giving up one's autonomy."<sup>58</sup> Thus, while such a promise would be binding, it would be impermissible to make that promise in the first place.<sup>59</sup> He also suggests that it makes no difference whether the promise is tacit or explicit.<sup>60</sup>

Thus, liberals cannot deny that autonomy is important without leaving behind liberalism itself. If they are to show that their states have legitimate authority, they must find a way of *reconciling* authority and autonomy. They need to find a way that an agent can fulfill their duties to deliberate and act on the basis of that deliberation, yet submit to the law just because it is the law. This challenge forms the basis for the rest of the thesis.

## 8 Conclusion

In *In Defense of Anarchism*, Wolff presents an argument that legitimate authority and autonomy are in conflict with one another. I have suggested that the crux of this argument concerns the way that we must deliberate and act as autonomous agents. An obedient agent fails to adequately act on the results of their deliberation, because they (a) exclude some of the considerations (especially moral considerations) that bear on what they should do, and (b) they treat the fact that something is commanded as reason to do it. I defended this argument against the charge that it leads to a *reductio ad absurdum* regarding promises and decisions. Finally, I pointed out that Wolff's argument is a vexing problem for liberals in particular, because they must accept Wolff's premise that autonomy has preeminent importance. In the next chapter, I consider the first of several ways that a liberal might attempt to reconcile authority and autonomy: moralism.

---

<sup>58</sup> Wolff, *In Defense of Anarchism*, 41.

<sup>59</sup> For similar reasons, most would say that we have a moral duty to avoid doing things that might lead us to slip, over time, into a state of *actual* non-responsibility. The most obvious example is taking impairing drugs – if someone consciously puts themselves in a position where they are incapable of free or rational decision-making (e.g. someone who drinks alcohol and then drives), we will likely negatively evaluate their decision to put themselves in that position. Likewise, it is possible for someone to culpably put themselves in a situation of ignorance, one in which they are genuinely no longer responsible for their actions, but that would be impermissible.

<sup>60</sup> Wolff, *In Defense of Anarchism*, 41n9. In the same footnote, Wolff also rejects hypothetical forms of consent entirely, suggesting that “It should be clear that a sophisticated interpretation of [hypothetical consent] will not do, if one wishes to found majority rule on the promise contained in the contract. A promise is an act, not the mere expression or summation of an existing obligation. It creates a new obligation where none existed before.”

## Chapter 3 Moralism

### 1 Introduction

As I argued in the last chapter, liberals hold that it is important for people to make up their own minds about their own lives, so they cannot answer Wolff's challenge by giving up on autonomy. Liberals must instead find a way to *reconcile* authority and autonomy, at least if they are to hold onto legitimate authority at all. The first attempt to reconcile them invokes what I will call *moralism* about political authority.

Many liberals (though not all) see the principles of liberalism as derived from a set of substantive moral values, like fairness, justice or even autonomy itself. I begin by describing how many of these liberals, including John Rawls in his earliest work on political obligation, have used some of these values to argue for the existence of a *prima facie* duty to obey the law – a duty to obey the law as long as it does not conflict with other moral values. To borrow a term from Bernard Williams, we might say that they are *moralists* about a *prima facie* duty to obey the law.<sup>61</sup>

While many liberal moralists are comfortable with a *prima facie* duty to obey the law, very few of them have used the same strategy to argue for the more demanding conclusion that the state has legitimate authority, which would imply that its citizens must treat laws as content-independent, exclusionary reasons. At first glance, this seems like a curious omission, because moralism about political authority would apparently offer a way of reconciling authority and autonomy. As we have seen, Wolff thinks that our duty to be autonomous is a duty to act on our moral duties, as we have understand them to be through our deliberation. But if an agent can

---

<sup>61</sup> I borrow the term “moralism” from Bernard Williams, but differ in usage from him. For Williams, the point of introducing “moralism” is to draw a contrast with “realism” in political philosophy. Rejecting “moralist” views that tie political theory too closely to moral theory, Williams suggests that liberal political theory should instead “shape its account of itself more realistically to what is plitudinously politics”. (Williams, “Realism and Moralism” in *In The Beginning was the Deed*, 12) For Williams, what is distinctive about political thought is that it is orientated around the “first political question” of “securing order, protection, safety, trust, and the conditions of cooperation”. (Williams, “Realism and Moralism” in *In The Beginning was the Deed*, 3)

plausibly reason that their moral duty *is* to obey the state, because such obedience is required by fair play or some other liberal value, then there would be no tension between legitimate authority and autonomy after all.

I then suggest an explanation for why liberals have not pursued this strategy. While moralism provides an adequate explanation for why citizens have a moral requirement to treat laws as content-independent reasons, it breaks down in trying to show that they should treat laws as *exclusionary* reasons. At most, moralism shows that citizens are required to treat commands or laws as ordinary moral reasons that sometimes outweigh competing moral commitments. But moralism does not show how laws or commands make those competing moral commitments cease to matter in the way that is characteristic of exclusion about how to act. Thus, liberal values cannot ground legitimate authority on their own after all. This failure is instructive. Moralism about political authority, at least in the simple form presented in this chapter, is a useful foil for understanding why the more complex explanations of political authority which I will discuss later in this thesis take the form that they do.

## 2 Moralism about the *Prima facie* Duty to Obey the Law

Many liberals see the principles of liberalism, including limits on state power and the protection of certain fundamental rights, as directly derived from a set of moral values, like fairness, justice or even autonomy itself. Some liberal philosophers use these same moral values to also argue that citizens have political obligations. Their target is often a *prima facie* duty to obey the law. A *prima facie* duty to obey the law is a duty to obey the law when it does not contradict other moral duties. To say that there is a *prima facie* duty to obey the law is a weaker claim than to say that the state has legitimate authority in Wolff's sense. Unlike a merely *prima facie* duty to obey the law, legitimate authority is often binding even when it contradicts other moral duties.<sup>62</sup>

---

<sup>62</sup> Though as I've suggested earlier, legitimate authority does not require that every command be binding, nor that every moral duty be excluded. So legitimate authority is, in terms of its demandingness, somewhere between a *prima facie* duty to obey the law and an absolute, exceptionless duty to obey the law.

Rawls's early fair play theory of political obligation is a good example of moralism about the *prima facie* duty to obey the law. In short, the duty of fair play is a duty to contribute to a just scheme of benefits and burdens from which one has accepted benefits.<sup>63</sup> We can most readily appreciate the force of the principle of fair play with examples drawn from outside of political philosophy. For example, the principle of fair play tells us that we should contribute to a potluck among a small group of friends if we eat from that potluck. A potluck is a mutually beneficial and just scheme of cooperation – there will be enough food for everyone if everyone puts in a dish, everyone benefits from having greater selection of dishes, and it is unlikely that there will be enough food to go around if not everyone puts in their fair share. It seems intuitively plausible that given this set of circumstances, if one takes from the potluck, one has a moral duty to likewise contribute to the potluck.

Rawls believes that this principle can also explain why members of a society have a *prima facie* duty to obey the law. If other members of a society are contributing to society through their obedience, and some person is benefiting from their obedience, that person should obey the law as well. This implies, for every law in a society, that the person has a moral reason of fairness to comply with it, one that will outweigh their non-moral reasons, but not always their moral reasons.

Borrowing a term from Bernard Williams, I will call this general strategy “moralism”.<sup>64</sup> In general, moralism is a very common position within the debate over the *prima facie* duty to obey the law. While no one in that debate explicitly calls themselves a moralist, the presence of moralism can sometimes be seen in how some of the figures in this debate describe the theoretical commitments behind their positions.<sup>65</sup> For example, A.J. Simmons, in one of the foundational surveys of political obligation, describes the need of a theory of political obligation

---

<sup>63</sup> Rawls, “Legal Obligation and the Duty of Fair Play” in *John Rawls: Collected Papers*, 122.

<sup>64</sup> In footnote 61, I explain how my usage of “moralism” originates in, but differs from, Williams’s usage.

<sup>65</sup> Note that many people in the debate over political legitimacy (rather than legitimate political authority) *do* explicitly call themselves moralists.

to explain the “special moral bonds” between a state and its members.<sup>66</sup> (Indeed, the moralism is written in the title of his book: *Moral Principles and Political Obligations*!) Likewise, M.L.E. Smith presents the main question of the debate over political obligation in the following terms: “Is the moral relation of any government to its citizens such that they have a *prima facie* obligation to do certain things merely because they are legally required to do so?”<sup>67</sup>

### 3 Moralism about Political Authority

In the previous section, I noted that some liberals attempt to derive a *prima facie* duty to obey the law from moral values like fairness. I will not take a stance on whether those attempts succeed.<sup>68</sup> Instead, I want to consider whether the liberal could use the same strategy to argue for the more demanding conclusion that citizens have a moral requirement to treat the law as a content-independent, exclusionary reason. On this view, which we might call moralism about political authority, legitimate authority is derived from liberal moral values. Few (if any) philosophers writing on legitimate authority seem to have defended this view despite its apparent simplicity, but it is important to see why it fails in order to understand why the liberal faces such a difficult task in accounting for legitimate political authority.

We might divide moralist theories of political authority into two categories. The first (‘inferentialist’) kind of moralist seeks to derive legitimate political authority from other moral duties that already have independent force. For example, the fair play principle appears to be drawn from moral theory in a way that is continuous with the rest of moral theory, since it seems applicable to situations inside and outside of politics, and Rawls identifies it as a principle whose force can be added to the inherent moral force of “wrongs as such”.<sup>69</sup> The second (‘*sui generis*’)

---

<sup>66</sup> A.J. Simmons, *Moral Principles and Political Obligations*, 4.

<sup>67</sup> M.L.E. Smith, “Is there a *prima facie* duty to obey the law?,” 952.

<sup>68</sup> As M.L.E. Smith’s article “Is there a *prima facie* duty to obey the law?” makes clear, in addition to the general arguments against the *prima facie* duty to obey the law, there are also many powerful arguments against the individual attempts to show that there is a *prima facie* duty to obey the law.

<sup>69</sup> Rawls, “Legal Obligation and the Duty of Fair Play” in *John Rawls: Collected Papers*, 118.

kind of moralism treats political authority as a *sui generis* moral phenomenon, one that has its own independent moral force even though it is not related to any other moral duties. This raises the question: how can we tell that a normative view is a moral one, without appealing to the view's inferential connection with the rest of morality? One might think that our most compelling reasons for action are always moral, so that any view that says that we have reason to obey an authority counts as a moralist view. This is reflected in Robert Ladenson's suggestion that "Given what we mean by the moral point of view, a person cannot intelligibly regard himself as being required by morality not to do what he thinks is required of him by morality."<sup>70</sup> However, I do not think that we should assume that this is the case, especially since the challenge of authority is find grounds by which an agent can justifiably exclude moral considerations from how they ultimately act. Yet this does show that we need a clearer specification of what counts as a moral view. I will assume one such view: we might distinguish moral views from non-moral views by considering how those views are justified – is the view derived in the same manner as our other moral beliefs (e.g. through intuition, or reflective equilibrium)? If it is, then it is a moral view.<sup>71</sup>

## 4 A Solution to Wolff's Challenge?

The moralist about legitimate authority has an initially plausible explanation for why obedience to an authority does not violate any duties of autonomy. As I have said, they claim that the obligation to obey a legitimate political authority is derived from (or part of) moral theory. In short, a moralist can claim that they must obey the authority *because of*, rather than *despite*, their moral views. Michael S. Pritchard appears to attribute such a view to Kant when arguing (against Wolff) that authority is consistent with Kantian autonomy after all:

---

<sup>70</sup> Robert F. Ladenson, "Legitimate Authority," 338.

<sup>71</sup> We can see Ronald Dworkin wrestling with this question in "Lord Devlin and the Enforcement of Morals". But in that article Dworkin only succeeds in producing some necessary conditions for something being a moral view, like the fact that the view must not be a prejudice or mere parroting of another person's view.

What Kant objects to is making one's moral will subject to the will of another, that is, one's right to make up his own mind about what is right or wrong. But that does not mean for Kant that one cannot do what another tells him to do, doing it because he has been told to do it... Kant's notion of moral autonomy permits one to decide that the right thing for him to do in some specifically instance is what has been legally prescribed...<sup>72</sup>

Pritchard is describing an agent who would be fulfilling their moral duties, as they understand them, by obeying an authority. They might be treating commands as reasons, and they might be ignoring some of their other moral considerations, but they take themselves as having moral justification for doing so. The subject still acts on their moral principles, so there is no conflict between authority and autonomy after all. The moralist would say that Wolff was correct about the value of autonomy, correct that we must act on assessments of what we think is morally correct, but that he failed to account for the possibility that that assessment might be to exclude some of our reasons.

Note that this explanation can go awry for some moralist views, on account of some of the considerations raised in the last chapter about consent theory. Recall that Wolff concedes that someone who consents to state rule through a promise would have a morally binding obligation to obey the law. At the same time, though, Wolff thinks that it would be impermissible to consent to such authority, because such consent would be consent to forfeit one's own autonomy. The example of consent theory, then, shows that some kinds of moralism (those where the obligation to submit to the state's authority is the result of certain voluntary acts) do not provide an adequate solution to Wolff's challenge.<sup>73</sup> But while this is a problem for some forms of moralism, it is important to remember that not all versions of moralism will be premised on voluntary acts whose permissibility or impermissibility can be called into question. If someone

---

<sup>72</sup> Pritchard, "Wolff's Anarchism," 297. See also Sterba, "The Decline of Wolff's Anarchism," 213.

<sup>73</sup> Note that this point might endanger the main example of moralism that I am using in this chapter, which is fair play, because fair play also hinges on voluntary actions. However, fair play hinges on a voluntary action that is much harder to avoid than consent. It would be very difficult to avoid partaking in the resources of a society, and perhaps that unavoidability explains why it is still permissible. Alternatively, even if fair play suffers from the same problem as consent theory, note that there are still many other potential moralist views of political authority that do not hinge on voluntary actions, like those based on duties to avoid causing harm to others, or the need to promote morally valuable goals in general.

finds themselves under a moral obligation to obey the state, but they didn't intentionally bring about this obligation themselves, they haven't acted wrongly.

That is the rough outline of how the moralist might try to reconcile authority and autonomy. However, in order for this proposal to work, it must be the case that there actually can be a wholly moral justification for excluding moral considerations from one's deliberations about how to act. I will now argue that such a justification is incoherent.

## 5 The Failure of the Solution

Recall in chapter 1 that I argued that there is a secondary conflict and a primary conflict between autonomy and authority. The secondary conflict comes from the *content-independence* of commands: an autonomous person should not treat a person's commands as reasons, because those commands are content-independent. The primary conflict comes from the *exclusion* involved in commands: obedience requires someone to exclude their own moral evaluation of the circumstances from their judgment about how to act.

I accept that the moralist adequately resolves the secondary conflict. Consider the fair play theory of political authority, which says that members of a society are obligated to obey the law (or submit to the state's authority) because other members of their society have done the same, and because they have benefited from their submission. It seems reasonable that the obligation to be fair in our dealings with one another could supply a moral justification for treating the mere fact that something is law as a strong reason for action. For each law, one would be violating a duty of fairness if one were to break it. In a slogan, considerations of fair play "trickle down" to each of the individual laws. Thus, it would be plausible to regard oneself as having moral justification for treating commands as reasons for action, even moral reasons for action.<sup>74</sup>

---

<sup>74</sup> One might frame this argument in a slightly different way: now that considerations of fairness are present, the fact that something is commanded is no longer a genuinely content-independent consideration at all. The fact that action is commanded implies that it would be unfair to disobey the command – it has become part of the content of the action that is relevant to deciding what to do.

On the other hand, I deny that the moralist can adequately resolve the primary conflict. The problem is that the moralist justification would seem to leave the force of any competing moral duties intact. In the fair play theory of political authority, for example, I grant that fairness gives us a further, moral reason in favour of acting on the content of the law, for much the same reason that a *prima facie* duty to obey the law might also be grounded on fairness. But any contrary moral commitments, like those that come from the evilness of the law in question, do not seem to be *excluded* by the existence of this duty derived from fairness. All fair play has done is added another moral duty to the mix. It is hard to see how someone in a fair play situation could be required to not even consider the other moral aspects of the actions they are commanded to do, in the way that someone who faces a red traffic light is supposed to not act on their assessment of whether they see traffic approaching the intersection.

This is particularly clear we think about the substantive moral considerations that derive from fairness itself. Suppose, for example, that the state was distributing benefits and burdens in a society in an unfair manner. On the one hand, fairness tells me to obey the law insofar as others in my society are doing the same. On the other hand, fairness itself seems to tell me that I should actively disobey the law, because my obedience to the law is furthering unfairness. It is hard to see how considerations of fairness are going to give me permission to exclude the latter consideration from my deliberations about how to act. If my commitment to obey the law is derived from fairness, it seems that I cannot coherently disregard the value of fairness in the same context. Such incoherence would mean that I would not be fulfilling my duties as an autonomous agent to deliberate in an appropriate way. One might argue in response that these are different kinds of unfairness, and so there is no incoherence in prioritizing one over the other, but this would imply that the kind of fairness that grounds obedience is independent of substantive considerations of fairness, and it seems implausible that the two forms have nothing in common with one another.

We can build on the previous argument by considering the nature of moral duties themselves. Consider one plausible view of how moral duties affect practical reasoning.<sup>75</sup> On the one hand, moral duties seem to exclude non-moral considerations from action. If I have a moral duty to keep a promise, or a moral duty not to harm others, this doesn't simply outweigh my personal prudential reasons, rather, it excludes them altogether. On the other hand, moral duties do not exclude one another. If my moral duty to keep a promise and moral duty not to harm others somehow come into conflict with one another, they do not exclude one another, but instead one outweighs the other.

## 6 Conclusion

I have argued that moral considerations do not explain why we can exclude other moral considerations from our deliberations, and so moralism does not address Wolff's challenge. Liberal moralists must draw the line at the *prima facie* duty to obey the law, and cannot go on to endorse legitimate political authority without appealing to some other argument.

The failure of moralism to solve Wolff's problem helps to clarify the nature of the challenge that faces liberals. They must make sense of why a citizen must exclude moral considerations from action. That is, they must find a standpoint from which it can make sense for an agent to not act on moral considerations that appear relevant to how to act. But in the coming chapters, I will argue that such a standpoint has proven elusive, at least in the leading liberal theories. In particular, I will argue that justice as fairness (John Rawls), the principles of practical reason (Joseph Raz), and the second-person standpoint (Stephen Darwall) do not provide this standpoint. This leaves liberals without an adequate theory of legitimate political authority.

---

<sup>75</sup> Raz, *Practical Reason and Norms*, 76. For further discussion, see Perry, "Second-Order Reasons, Uncertainty, and Legal Theory". W.D. Ross's discussion of *prima facie* duties in *The Right and the Good* also supports this picture of moral duties.

## Chapter 4 Justice as Fairness

### 1 Introduction

In this chapter, I argue that John Rawls's "justice as fairness" does not satisfactorily address Wolff's problem. In short, justice as fairness is a system of liberal political morality that Rawls progressively develops in his works *A Theory of Justice* and *Political Liberalism*. It is a set of principles for a society and its members that is derived from the original position, a thought experiment in which agents choose principles for themselves and their institutions while lacking certain information about the position they will occupy in that society.

I begin by presenting the theory of political authority and political obligation found in *A Theory of Justice*, which involves a complex, multi-stage application of the original position. As a social contract theory, it might initially appear to be a form of consent theory, and if so, it would be unacceptable to Wolff for that reason. However, Rawls uses the idea of a social contract as a device for revealing the "fair terms of cooperation" and does not rely on the binding force of consent to explain the authority of the revealed principles. Instead, I think we should understand justice as fairness as a very intricate moralist theory, one that has two initial advantages over simpler moralist theories. First, justice as fairness can be interpreted in Kantian terms, and thus Rawls clearly expects that his theory of political authority is consistent with Kantian autonomy. Surely if any moralist theory can reconcile authority and autonomy, it would be a Kantian one. Second, Rawls has a novel, two-step strategy for explaining why citizens must exclude their own moral judgments when obeying the law. At step one, liberal moral commitments are given by the original position. At step two, the choices made by agents in the original position determine the principles that will govern their society's institutions and individuals, including principles of political authority. This frees Rawls from the strictures of simple moralism, allowing him to argue that people in the original position would choose to be bound by principles of political obligation that require authoritative exclusion as a way of ensuring the efficacy and stability of the principles of justice for their society's institutions.

Despite these two advantages over simpler forms of moralism, I argue that justice as fairness does not successfully reconcile authority and autonomy. While the original position is a plausible way of construing autonomous reasoning, I argue that Rawls is incorrect that people in the original position would choose to be bound by a principle that requires them to not act on certain moral considerations that they deem relevant. In short, they would choose to leave it open to themselves to judge whether the loss of stability from disobeying an unjust law outweighs the injustice of the law, rather than excluding any of these considerations. Thus, the strategy fails at its second step.

In *Political Liberalism*, Rawls instead presents justice as fairness in a different light. Rather than offering it as a correct moral view, or as a competitor to other moral views, he now proposes it as a basis for an “overlapping consensus” between reasonable moral views (or “comprehensive doctrines”). Rawls now hopes to show that justice as fairness, presumably including its principles of political authority, is acceptable to each reasonable comprehensive doctrine in circulation within a society. I argue that this new presentation of justice of fairness simply pushes the problem of explaining authoritative exclusion onto the individual comprehensive doctrines. That is, we still need an explanation of how any comprehensive doctrine (e.g. justice as fairness or utilitarianism) could require its holder to not act on some moral considerations.

## 2 *A Theory of Justice: Overview*

In *A Theory of Justice*, Rawls defends a version of social contract theory. Unlike many earlier social contract theories, like those of Hobbes and Locke, which use the idea of a social contract to directly explain the legitimacy or authority of government, Rawls uses the social contract as a device for revealing the fair terms of cooperation between members of a society. To determine an appropriately fair set of principles to govern a society and its institutions, Rawls suggests we must imagine what contractors would choose in the “original position”.<sup>76</sup> In this imagined

---

<sup>76</sup> Rawls, *A Theory of Justice*, 3.

situation, “rational and mutually disinterested” agents choose the principles that would apply to themselves and their institutions.<sup>77</sup> They make this choice from behind a “veil of ignorance”, so they would not know their place in society, class position, nor their share of natural assets and abilities.<sup>78</sup>

Rawls uses the original position to show that which principles should govern the basic structure or institutions of a society. He concludes that agents in the original position would find two principles to be most choiceworthy (at least in comparison with the alternatives he considers). These principles are now very well-known. The first principle requires that there should be an equal distribution of certain basic liberties, while the second principle requires that social and economic inequalities should only exist when they are for the benefit for everyone, particularly the worst-off members of a society, and that advantageous positions be attached to offices that are in principle open to all.<sup>79</sup>

As a check on the original position, Rawls thinks that it is important to verify that the chosen principles would be practicable and stable by the people who will live by them. In evaluating the practicability of the two principles of justice, he argues that they avoid the “strains of commitment”, particularly when they are compared with utilitarianism, and so they meet one necessary condition to be practicable.<sup>80</sup> Rawls uses “stability” in a slightly more technical way, writing that, “An important feature of a conception of justice is that it should generate its own support. Its principles should be such that when they are embedded in the basic structure of society men tend to acquire the corresponding sense of justice and develop a desire to act in accordance with its principles. In this case a conception of justice is stable.”<sup>81</sup> Elsewhere, he describes it in slightly different terms: stable principles “must be more or less regularly complied with and its basic rules willingly acted upon; and when infractions occur, stabilizing forces

---

<sup>77</sup> Rawls, *A Theory of Justice*, 12.

<sup>78</sup> Rawls, *A Theory of Justice*, 10.

<sup>79</sup> Rawls, *A Theory of Justice*, 13.

<sup>80</sup> Rawls, *A Theory of Justice*, 153.

<sup>81</sup> Rawls, *A Theory of Justice*, 119. See also 393-4.

should exist that prevent further violations and tend to restore the arrangement.”<sup>82</sup> Later in *A Theory of Justice*, Rawls provides an extended account of the moral psychology of human beings that he uses to argue that his two principles are stable.<sup>83</sup>

### 3 *A Theory of Justice: Political Authority*

Having explained the original position and used it to generate principles of justice for a society’s institutions, Rawls next applies it to the more concrete question of how to design and implement a society’s institutions. First, he uses it to derive a society’s basic constitutional procedure – how to translate the abstract principles of justice into concrete laws. Second, he uses it to derive the principles for the individual conduct of a society’s members. These stages of his analysis contain his theory of political authority and political obligation, which I will present in detail in this section.

The next stage of Rawls’s analysis is to use the original position to determine a society’s “constitutional principle” – the principle that dictates which laws should be passed, having already fixed the principles of justice for institutions. A constitutional principle is necessary because even in a perfectly well-ordered society, there is a gap between the principles of justice and the particular laws that are intended to embody them. There are no preordained rules about how to write legislation that embodies the two principles of justice for institutions. (For this reason, Rawls thinks the constitutional principle is still a matter for “ideal” theory. Constitutional questions will need to be resolved even if we assume that people in a society observe the same set of principles.<sup>84</sup>) Rawls explicitly considers the possibility that there might be a single, authoritative interpretation of the principles of justice, but rejects it. He writes, “In a democratic society, then, it is recognized that each citizen is responsible for his interpretation of the principles of justice and for his conduct in light of them. There can be no legal or socially

---

<sup>82</sup> Rawls, *A Theory of Justice*, 6.

<sup>83</sup> Rawls, *A Theory of Justice*, chapter 8.

<sup>84</sup> Rawls, *A Theory of Justice*, 315.

approved rendering of these principles that we are always morally bound to accept, not even when it is given by a supreme court or legislature.”<sup>85</sup>

Even in a well-ordered society where the principles of justice are accepted by all, then, it is still necessary to determine a decision procedure for translating the principles into concrete legislation. This procedure is determined by a further iteration of the original position, one with access to more information than before, including the “relevant general facts about their society”, including its “natural circumstances”, “level of economic advance” and “political culture”.<sup>86</sup>

Rawls suggests two desiderata that a constitutional procedure must satisfy.<sup>87</sup> First, the procedure must itself satisfy the principles of justice, particularly the first principle of justice and the first part of the second principle, so that everyone will have equal political rights and access to its offices. Second, Rawls suggests that people in the original position will choose the decision procedure that is most likely to result in just legislation, using the standard of justice found in the two principles of justice. By these criteria, Rawls thinks that some form of majority rule (perhaps one constrained by a constitution) will best satisfy the desiderata mentioned earlier. He notes that majority rule is compatible with equal liberty, avoids the violations of equality that accompany minority rule, and so is the “best available way of insuring just and effective legislation.”<sup>88</sup>

The final stage of Rawls’s analysis is to use the original position to determine principles for individual conduct.<sup>89</sup> Rawls divides these principles into two categories. The first category concerns natural obligations, which are created when someone undertakes a voluntary action. Rawls suggests that all obligations, including obligations of promise-keeping and fidelity, are related to a single obligation of fairness or “fair play”. The second category concerns natural

---

<sup>85</sup> Rawls, *A Theory of Justice*, 342.

<sup>86</sup> Rawls, *A Theory of Justice*, 172.

<sup>87</sup> It is somewhat unclear whether these desiderata would be determined within the original position itself, or are part of the original position.

<sup>88</sup> Rawls, *A Theory of Justice*, 313.

<sup>89</sup> Note that in Rawls’s framework, there appears to be no way for individuals to act on the principles of institutional justice directly for themselves. It is always mediated by principles for individual conduct. An individual’s sense of justice, in turn, provides the desire for complying with these principles. (Rawls, *A Theory of Justice*, 128)

duties, which bind people independently of their voluntary actions, including the duty of mutual aid, the duty of mutual respect, and the duty that will turn out to underwrite political obligation: the duty to support and further just institutions.

Rawls treats the principles for individual conduct that are derived from the original position as strongly *complementary* to the principles for institutional justice. In other words, Rawls thinks that it is only possible to determine the principles of individual conduct once the content of the principles of justice for institutions (and the constitutional principle) have been settled.<sup>90</sup> This complementarity has an important implication for how they are chosen in the original position: Rawls appears to think that agents would choose between principles for individual conduct solely on the basis of how well they complement the principles of justice. They would not take into consideration how well the principle might accord with individual moral views or support individual happiness, only their effect on the justice of society as a whole.<sup>91</sup> This comes out most strikingly in his justification of the principle of mutual aid, which imposes a duty to help others where it is reasonable to do so. Ordinarily, we would expect a principle of mutual aid to be grounded in terms of considerations of individual interest, but Rawls instead talks about its effect on “the sense of confidence and trust in other men’s good intentions and the knowledge that they are there if we need them.”<sup>92</sup> He seems to be saying here that the point of the duty of mutual aid is to help stabilize the principles of justice. People in the original position would choose it not because they imagine that they might need help at some point, but because they want to live in a society where it is known that people help one another.

Assuming that the principles that best complement the principles of justice would be chosen, Rawls thinks that it is clear that the most complementary principle would be the one most tailored to the principles of institutional justice: a duty to “support and further just

---

<sup>90</sup> Rawls, *A Theory of Justice*, 95.

<sup>91</sup> Rawls suggests that considerations of “personal prudence are already given an appropriate weight within the full system of principles.” (Rawls, *A Theory of Justice*, 117)

<sup>92</sup> Rawls, *A Theory of Justice*, 298.

institutions”.<sup>93</sup> This duty has two parts. First, it requires us “to comply with and to do our share in just institutions when they exist and apply to us.”<sup>94</sup> Second, it requires us to “assist in the establishment of just arrangements when they do not exist, at least when this can be done with little cost to ourselves.”<sup>95</sup>

He also addresses a potential counterargument that would aim to show that one would not choose the duty of natural justice from the original position. In previous work, he had sought to ground political obligation in terms of fair play, apparently on intuitionistic grounds rather than through the original position. In short, the duty of fair play is a duty to contribute to a scheme of benefits and burdens from which one has accepted benefits.<sup>96</sup> Why wouldn’t people in the original position let a principle of fairness be the sole source of their political obligation? At first glance, even though it might not be as directly tailored to assisting just institutions, in one respect fairness is perhaps more attractive than a duty to support and further just institutions: a principle of fairness might appear to cohere better with the social contract’s underlying ideals of “free consent and the protection of liberty”, since it makes political obligation rest on a voluntary act.<sup>97</sup> However, Rawls thinks that this is irrelevant, because liberties are already secured through the principles for institutions in a well-ordered society.<sup>98</sup> Moreover, Rawls suggests that even in a well-ordered society, worries about assurance and stability can still arise. People in a well-ordered society are only imperfectly motivated, and so would have reason to worry that other members of their society will join them in obeying the law. For reasons of stability, it is necessary to find the “easiest and most direct” way of guaranteeing that citizens will support

---

<sup>93</sup> Rawls, *A Theory of Justice*, 293.

<sup>94</sup> Rawls, *A Theory of Justice*, 293.

<sup>95</sup> Rawls, *A Theory of Justice*, 293.

<sup>96</sup> Rawls, “Legal Obligation and the Duty of Fair Play” in *John Rawls: Collected Papers*, 122.

<sup>97</sup> Rawls, *A Theory of Justice*, 295. Note that if it were a form of consent theory, it would also fail as a solution to Wolff’s problem (just for different reasons).

<sup>98</sup> Rawls, *A Theory of Justice*, 295.

their institutions, and the principle of fairness is not sufficiently direct.<sup>99</sup> It only applies when they have voluntarily accepted the benefits provided by those institutions. Citizens will not believe themselves to be bound by the duty of fair play if they believe they did not voluntarily receive the benefits of their society, or if they believe that other citizens are not contributing their share. This in turn would force the state to use its “coercive powers” to achieve stability, which he describes as a “risk” to be avoided.<sup>100</sup> Thus, the principle of fairness is not completely discarded, and does provide some further reason to comply with the law, but it now plays only a subsidiary role in stabilizing the state’s institutions. Thus in a perfectly well-ordered society, one that is effectively regulated by the principles of justice, people would choose to be bound by a natural duty to support and further just institutions.

However, Rawls suggests that the case of unjust law raises special difficulties.<sup>101</sup> Unjust legislation is an inevitable product of the constitutional principle of majority rule – even when acting in good faith, both the democratic system and its individual actors will make mistakes that result in unjust legislation being enacted. In light of this, people in the original position would recognize that they would sometimes be confronted with laws that seem to them to be somewhat unjust. However, given the inevitability of these injustices, Rawls thinks that we would see that it is nevertheless necessary to submit to them sometimes. So there must be some principle that sets out when we have an obligation to obey the law even when it appears unjust to us. He supplies this principle when he writes, “If the law actually voted [into law] is, so far as one can ascertain, within the range of those that could reasonably be favored by rational legislators conscientiously trying to follow the principles of justice, then the decision of the majority is practically authoritative, though not definitive.”<sup>102</sup> To show that a law is unjust, a citizen must convince themselves that it could not be adopted by legislators guided by the two principles of justice.

---

<sup>99</sup> Rawls, *A Theory of Justice*, 295.

<sup>100</sup> Rawls, *A Theory of Justice*, 297. Rawls does not directly say why the state using its coercive powers to achieve stability is a “risk”; he takes it as evident that it is undesirable from the original position.

<sup>101</sup> As Rawls notes, one complication with this argument is that the society may no longer be a well-ordered society, so the assumptions of the original position may no longer hold.

<sup>102</sup> Rawls, *A Theory of Justice*, 318. Rawls presents this principle in the context of an ideal, well-ordered society, but it appears to extend to the case of unjust law as well.

## 4 *A Theory of Justice: A Solution to Wolff's Challenge?*

How would justice as fairness, as presented in *A Theory of Justice*, respond to Wolff's challenge, and reconcile authority and autonomy with one another? Recall in chapter 1 that I argued that there are two conflicts between autonomy and authority. The secondary conflict concerns content-independence of commands, and comes from an autonomous person's moral obligation to not treat a person's commands as reasons. The primary conflict concerns the exclusionary nature of commands, and comes an autonomous person's obligation not to exclude their moral evaluation of the circumstances in deciding how to act. Both of these were problematic because of the apparent incoherence of treating a moral consideration as excluding another moral consideration. Can justice as fairness make sense of why an obedient person might exclude moral considerations in deciding how to act?

It is important to first note that although justice as fairness is a version of social contract theory, Rawls is not saying that authority is binding because it is accepted. He is not defending a form of consent theory. Instead, he thinks that the original position is a device for revealing the content of the fair terms of cooperation. It is not the choice itself, but what the choice reveals about the fair terms of cooperation, that explains the bindingness of the principles chosen in the original position. As Ronald Dworkin puts the point, "it must be a device for calling attention to some independent argument for the fairness of the two principles – an argument that does not rest on the false premise that a hypothetical contract has some pale binding force."<sup>103</sup>

Instead, the social contract describes how an autonomous person should deliberate. In fact, Rawls thinks that the original position can be given a "Kantian interpretation", since it "may be viewed" as a "procedural interpretation of Kant's conception of autonomy and the categorical imperative within the framework of an empirical theory".<sup>104</sup> As he writes, "Kant held, I believe, that a person is acting autonomously when the principles of his action are chosen by him as the

---

<sup>103</sup> Dworkin, "The Original Position," 520.

<sup>104</sup> Rawls, *A Theory of Justice*, 226.

most adequate possible expression of his nature as a free and equal rational being.”<sup>105</sup> So Rawls agrees with Wolff that it is necessary to respect autonomy in the Kantian sense. Indeed, Rawls’s description of an agent deliberating about the justice of a law calls to mind Wolff’s argument in *In Defense in Anarchism*:

There are parallels with the common understandings and conclusions reached in the sciences. Here, too, everyone is autonomous yet responsible. We are to assess theories and hypotheses in the light of the evidence by publicly recognized principles. It is true that there are authoritative works, but these sum up the consensus of many persons each deciding for himself.<sup>106</sup>

That said, there are differences between how Rawls and Kant treat autonomous deliberation. For Wolff, autonomous deliberation is very open-ended. He appears to think that an agent must take into account in their deliberations almost any consideration that they see as morally relevant. In contrast, Rawls imposes more mandatory structure onto the deliberation of autonomous agents. For example, he notes that the veil of ignorance reflects a way of blocking out “heteronomous” considerations from deliberation.<sup>107</sup> That said, Rawls says nothing at all about whether it is permissible or impermissible to treat commands as reasons, or to exclude certain considerations from one’s deliberations. Thus, while there are differences between Rawls and Wolff’s views of autonomous deliberation, the original position at least appears to be *consistent* with Wolff’s view of how autonomous agents must deliberate. That is, if an agent decided to choose their moral principles through the original position, I do not think that Wolff would not have a quarrel with them.

If we are to understand the original position (and justice as fairness more generally) as addressing Wolff’s challenge, it should be understood as a kind of moralism.<sup>108</sup> Rawls would say

---

<sup>105</sup> Rawls, *A Theory of Justice*, 222.

<sup>106</sup> Rawls, *A Theory of Justice*, 341.

<sup>107</sup> Rawls, *A Theory of Justice*, 222.

<sup>108</sup> To verify that Rawls is a moralist about political authority, it is also necessary to show that justice as fairness really is a moral theory. I think that is fairly clear from his writing. He presents justice as fairness as a competitor to

that the autonomous agent will ultimately see that their moral duty is to obey the state, and so there is no conflict between their obedience and their autonomy. But unlike the theories of moralism discussed in the last chapter, Rawls's moralism has two steps. First, the fair terms of cooperation, which reflect justice for a society and its institutions, are set by the original position. Second, within the original position, the specific principles for a society and its members are chosen. The second step details how Rawls might try to resolve the primary and secondary conflicts between authority and autonomy. An agent, taking the original position as the source of their moral principles, will find that those principles include a requirement to exclude their assessment that a law is somewhat unjust from their deliberation, and just obey the law.<sup>109</sup>

One might wonder whether Rawls's view of political authority in *A Theory of Justice* is genuinely exclusionary. Perhaps he is merely aiming to show a weaker conclusion, like the existence of a *prima facie* duty to obey the law. However, I think that his position here is exclusionary. For example, he writes that "It suffices to note that while citizens normally submit their conduct to democratic authority, that is, recognize the outcome of a vote as establishing a binding rule, other things equal, they do not submit their judgment to it."<sup>110</sup> This sounds remarkably like the observation that Raz made in response to Hart: people who submit to

---

other moral theories, like intuitionism, perfectionism and utilitarianism, which suggests that he intends for it to be a moral theory on par with them. The original position is also supposed to play a similar role in practical reasoning as a moral theory. He suggests that the requirements of a conception of right are "decisive", "override the demands of law and custom" as well as "considerations of prudence and self-interest". (Rawls, *A Theory of Justice*, 116) They are "final in that when the course of practical reasoning it defines has reached its conclusion, the question is settled. The claims of existing social arrangements and of self-interest have been duly allowed for. We cannot at the end count them a second time because we do not like the result." (Rawls, *A Theory of Justice*, 117)

That said, there is a thought that Rawls occasionally pursues in *A Theory of Justice* about the original position, one that has the potential to lift it beyond moralism. At one point, he describes the original position as "a common point of view from which their claims may be adjudicated". (Rawls, *A Theory of Justice*, 4) Likewise, he describes it as a way of "reach[ing] some measure of understanding and agreement which goes beyond a mere de facto resolution of competing interests and a reliance on existing conventions and established expectations..." (Rawls, *A Theory of Justice*, 31) These hint at an alternative reading of justice as fairness: perhaps the original position might be offered not as a moral theory in itself, intended to rival utilitarianism or intuitionism, but instead as a way of arbitrating between different people who endorse moral theories, who need not give up those moral theories, but instead must defer to the principles when they face conflict with others.

<sup>109</sup> As we have seen, he thinks that this extends to only moderate injustices, not severe injustices.

<sup>110</sup> Rawls, *A Theory of Justice*, 314.

authorities retain the freedom to judge matters for themselves, but are expected to exclude certain reasons when acting.<sup>111</sup>

Before evaluating this argument, it is illuminating to compare it with the one-step moralist strategy for addressing the primary conflict between authority and autonomy. To use the example of the fair play theorist, one could take a single moral principle (fair play) and try to make sense of how moral fairness could give someone reason to not act their other moral reasons. Unlike the fair play theorist, Rawls does not need to make sense of how one could infer the exclusionary reason directly from other moral duties. Instead, the fact that an agent would choose an exclusionary principle from within the original position is enough to make sense of why an agent might be bound by a principle that required such exclusion. Thus, because he defends a two-step version of moralism, Rawls does not face the same obstacle as the other moralists in making sense of authoritative exclusion.

## 5 *A Theory of Justice: The Failure of the Solution*

I will now argue that justice as fairness is not successful in reconciling authority and autonomy. Despite its Kantian origins, and even though it is a two-step process, justice as fairness is unable to resolve the primary conflict between authority and autonomy.

Rawls's argument that authoritative exclusion can be reconciled with autonomy, which I discussed in the previous section, might be summarized in the following way:

- (1) As long as an agent is acting on a principle that would be chosen in the original position, they are acting autonomously.
- (2) In the original position, agents would choose a principle requiring them to support just institutions, which entails that they must not act on certain moral considerations (that is, they must exclude them from their deliberations in how to act).

---

<sup>111</sup> Moreover, even if Rawls did not intend to defend an explicitly exclusionary principle, I think it is still worth asking to ask whether the original position *could* justify such a principle.

- (3) In acting on a duty to support just institutions, and not acting on certain moral considerations, agents act autonomously.

Moving on to evaluate the argument, I think that premise (1) is plausible for reasons already mentioned in the previous section. Rawls thinks that the original position is a device for revealing information about the fair terms of cooperation. Seen in this light, the original position is one way of deliberating about moral questions that fulfills Wolff's duties of autonomy. It is not the only way that someone can deliberate about moral questions and still fulfill their obligations to deliberate and act on the results of that deliberation, but it is certainly one way that they can do so.

On the other hand, premise (2) is mistaken. I will argue that agents in the original position would not accept a principle that requires or even permits them to exclude certain considerations from their deliberations in how to act. To frame this argument, it will be helpful to compare two principles:

(A): Rawls's *exclusionary* principle, which again says that "If the law actually voted [into law] is, so far as one can ascertain, within the range of those that could reasonably be favored by rational legislators conscientiously trying to follow the principles of justice, then the decision of the majority is practically authoritative, though not definitive."<sup>112</sup>

(B): An *anarchist* principle that permits agents to weigh the relative costs of obedience and disobedience and decide on that basis whether to obey or disobey the law.

Rawls does not explicitly consider (B), but as we saw earlier, he does offer a stand-alone argument that agents would choose (A).<sup>113</sup> He argues that the adoption of a constitutional principle that permits some degree of injustice is unavoidable. Agents in the original position

---

<sup>112</sup> Rawls, *A Theory of Justice*, 318. Again, I grant that this is a somewhat restrained exclusionary principle, but I think it is an exclusionary principle nevertheless.

<sup>113</sup> He frames the need for such an argument in a way that at first glance sounds quite sensitive to Wolff's concerns, "One might ask: how is it possible that when we are free and still without chains, we can rationally accept a procedure that may decide against our own opinion and give effect to that of others?" (Rawls, *A Theory of Justice*, 311)

appreciate that such a constitutional principle will produce legislation that inevitably clashes with their own sense of justice, and they accept such a principle nevertheless.

I agree that Rawls has shown that some form of constitutional principle is necessary. However, it is unclear why that implies anything about an individual's obligations. In defense of his principle of political obligation, he writes that "The contract doctrine naturally leads us to wonder how we could ever consent to a constitutional rule that would require us to comply with laws that we think are unjust."<sup>114</sup> But as Rawls has made clear, the constitutional principle and principles of individual obligation are separate stages of the application of the original position. Majority rule seems consistent with many different principles of individual duty or obligation, so why couldn't parties accept majority rule as their procedure for law-making, but deny that it should play any exclusionary role in their individual reasoning?

What seems to be doing the actual work in Rawls's argument for (A) is not the constitutional procedure itself, but instead his original argument for the duty to support and further just institutions, that widespread obedience is necessary to ensure efficacy and stability. As he notes, "...the parties accept the risks of suffering the defects of one another's knowledge and sense of justice in order to gain the advantages of an effective legislative procedure."<sup>115</sup> As a stand-alone argument for (A), this argument is initially plausible, but when we reconsider it with (B) as an apparent alternative, we can see that it is not as strong as it first appears to be. In a society whose members follow (B), there is still one set of laws to follow, produced by the single constitutional principle, but every citizen will have to make up their own mind about whether to obey the laws, depending on how they weigh, from their point of view, the injustice of the law against the destabilizing force of disobedience. It will be analogous, in a way, to Wolff's lifeboat example, where the passenger must take into account the captain's instructions, but decide for themselves. It's unclear that a model like this would make the principles of justice less efficacious, because citizens would still take into account the instability caused by their disobedience. Moreover, when we think of broader questions of stability, like the adoption and

---

<sup>114</sup> Rawls, *A Theory of Justice*, 311.

<sup>115</sup> Rawls, *A Theory of Justice*, 312.

stability of the principles of justice in the long term, (A) is not superior to (B). This is particularly true regarding civil disobedience, the stabilizing effects of which Rawls praises elsewhere: “By resisting injustice within the limits of fidelity to law, it serves to inhibit departures from justice and to correct them when they occur. A general disposition to engage in justified civil disobedience introduces stability into a well-ordered society, or one that is nearly just.”<sup>116</sup>

So far, in arguing that (A) is not superior to (B), I have granted Rawls’s assumption that arguments within the original position for principles for individual conduct must invoke their relationship to the principles of justice. Moreover, I have granted his assumption that the original position takes a very austere view of the moral dispositions of the people for whom principles are being chosen – people in the original position only know that they have a conception of the good, a rational plan for achieving that good, and a sense of justice, and not that they are particularly interested in each other’s ends. To an extent, I grant that he has a principled reason for making these assumptions about the original position. He does not want to assume going into the original position that people have certain duties towards one another.<sup>117</sup> Instead, he wants to derive all duties from the original position itself. However, by relying on this set of assumptions, Rawls prevents by fiat the kind of reasoning that Wolff invokes in defense of autonomy from entering into the original position. For Wolff, the duty of autonomy is not in place to assist agents more effectively enact just principles in their societies. Instead, the duty of autonomy is intended to be a direct consequence of their agency. One might imagine that if agents were freed from Rawls’s assumptions, they would choose to impose this duty on themselves. This suggests a further argument for (B). Surely agents themselves have a deep concern with whether they will wind up acting unjustly. Knowing that they will have a sense of injustice, they would not risk forcing themselves to act on injustice. This seems particularly true regarding considerations that seem to them to be morally important but cannot be understood in terms of the fair terms of cooperation, which they would want to stay at liberty to act upon, because they are not captured under the original logic of the principles of political obligation.

---

<sup>116</sup> Rawls, *A Theory of Justice*, 336.

<sup>117</sup> Rawls, *A Theory of Justice*, 111.

To sum up, I have argued that despite its Kantian affinities, justice as fairness does no better than ordinary moralism in reconciling authority and autonomy. To close my discussion of *A Theory of Justice*, and make way for my discussion of *Political Liberalism*, it will be helpful to make one transitional observation. There is one feature of justice as fairness as presented in *A Theory of Justice* that I think helps to more deeply explain its failure to explain the exclusionary force of political obligation. This is his assumption that society is well-ordered. As we have seen, Rawls suggests that in choosing principles in the original position, the contractors assume that the society they will occupy is “well-ordered” – one in which “everyone accepts and knows that the others accept the same principles of justice”.<sup>118</sup> Under that assumption, there is much less need for justice as fairness to have exclusionary force; it is already posited that it will be accepted by the members of the society for whom it is chosen. It is the rejection of this assumption that leads Rawls to go beyond moralism in his next work, *Political Liberalism*.

## 6 *Political Liberalism: Overview*

Soon after the publication of *A Theory of Justice*, Rawls came to realize that his assumption of well-orderedness was a great liability to his view. Participants in the original position choose principles for a society in which “everyone accepts and knows that the others accept the same principles of justice”.<sup>119</sup> In *Political Liberalism*, Rawls concedes it is unlikely that everyone will accept the same principles of justice in any society remotely like our own. Not only do people endorse very different principles of justice, but this variety is a “desirable output of the exercise of practical reason within free institutions.”<sup>120</sup>

---

<sup>118</sup> Rawls, *A Theory of Justice*, 4.

<sup>119</sup> Rawls, *A Theory of Justice*, 4.

<sup>120</sup> Rawls, *Political Liberalism*, 37.

The failure of this assumption raises two problems for justice as fairness.<sup>121</sup> The first is familiar. In *A Theory of Justice*, Rawls had insisted that any plausible set of principles of justice must not only be substantively choiceworthy but also stable – they “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.”<sup>122</sup> Yet it is very hard to see how justice as fairness could be stable for a society whose members endorse other principles of justice, since it is unlikely that they would willingly follow its prescriptions. Thus, it would not make sense to choose justice as fairness in the original position as a guiding principle for a non-ideal society. The second problem is new. If most members of a society do not endorse justice as fairness, it is hard to see how enforcing justice as fairness on them could be legitimate. Rawls describes his view of legitimacy in what he terms the “Liberal Principle of Legitimacy”:

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.<sup>123</sup>

Rawls’s overall goal in *Political Liberalism* is to find a sense in which a morally diverse society can find common political ground that is both legitimate and stable – perhaps, though not necessarily, on the basis of justice as fairness.

As part of his explanation of how this might be possible, he introduces three new concepts. The first is a *comprehensive doctrine*, a set of principles which “organizes and characterizes recognized values so that they are compatible with one another and express an intelligible view of the world” and “covers the major religious, philosophical and moral aspects

---

<sup>121</sup> In organizing Rawls’s changes between *Theory of Justice* and *Political Liberalism* as a response to these two problems, I follow Leif Winar, “John Rawls”.

<sup>122</sup> Rawls, *A Theory of Justice*, 6.

<sup>123</sup> Rawls, *Political Liberalism*, 137.

of human life in a more or less consistent and coherent manner.”<sup>124</sup> The theories of justice presented in *A Theory of Justice*, including justice as fairness as presented there, utilitarianism, intuitionism, as well as many other religious and secular doctrines, all count as comprehensive doctrines. Rawls grants that we live in a society whose members endorse many different comprehensive doctrines, and cannot be expected to change those doctrines. However he does expect that people would not seek to impose their comprehensive doctrines on others using force, particularly state force – such persons (and their doctrines) would count as “unreasonable”.<sup>125</sup>

The second concept is a *political conception of justice*. A political conception of justice has a narrower range than a comprehensive doctrine. It applies only to the “basic structure” of society: a society’s political, social and economic institutions.<sup>126</sup> At a minimum, it serves to express a view of what fair terms of cooperation might be – that “all who are engaged in cooperation and who do their part as the rules and procedures require, are to benefit in an appropriate way as assessed by a suitable benchmark of comparison”.<sup>127</sup> Rawls calls for restraint in choosing which matters should be addressed by political conceptions of justice – keeping them relatively few and urgent, and avoiding “divisive” issues that might “undermine the bases of social cooperation”.<sup>128</sup> Doing so will hopefully allow a society to “bypass religion and philosophy’s profoundest controversies”.<sup>129</sup> Most importantly, a political conception of justice is intended to be “freestanding” from individual comprehensive doctrines – it is not itself an individual comprehensive doctrine, nor derived from any of them, but instead is a “module” that “fits into and can be supported by various reasonable comprehensive doctrines that endure in the society regulated by it”.<sup>130</sup>

---

<sup>124</sup> Rawls, *Political Liberalism*, 59.

<sup>125</sup> Rawls, *Political Liberalism*, 60.

<sup>126</sup> Rawls, *Political Liberalism*, 11.

<sup>127</sup> Rawls, *Political Liberalism*, 16.

<sup>128</sup> Rawls, *Political Liberalism*, 157.

<sup>129</sup> Rawls, *Political Liberalism*, 152.

<sup>130</sup> Rawls, *Political Liberalism*, 12.

The third concept is an *overlapping consensus* regarding some political conception of justice. There is an overlapping consensus among the members of a society regarding some political conception of justice when each of them endorse that political conception of justice for moral reasons internal to those own comprehensive doctrines. As Rawls writes, such a conception “provides a publicly recognized point of view from which all citizens can examine before one another whether their political and social institutions are just. It enables them to do this by citing what are publicly recognized among them as valid and sufficient reasons singled out by that conception itself.”<sup>131</sup> For example, certain religious doctrines may endorse freedom of speech on the grounds that it shows appropriate respect for God’s endowment of human beings with reason, while utilitarians may endorse freedom of speech on the grounds that it leads to a happier, more prosperous society.

It is quite unlikely that any comprehensive doctrine, or any set of rules explicitly presented as a fragment of some comprehensive doctrine, could be a basis of such an overlapping consensus. Instead, Rawls suggests that we will most likely find an overlapping consensus in political conceptions of justice that are interpretations of certain fundamental ideas already present in the political culture of a society. In the case of a country with a political culture like the United States, with its constitution, bill of rights, and so on, an overlapping consensus will be based on an interpretation of ideas like liberty and equality and fairness inherent in the political culture. By rooting the political conception of justice in the fundamental ideas of a society’s political culture, Rawls hopes that it is reasonable to expect that it could command the acceptance of the members of that society. Moreover, Rawls hopes that once an overlapping consensus is found, it would persist over time. It would be more likely to persist than a “modus vivendi”, that is, a “social consensus founded on self- or group interests, or on the outcome of political bargaining”.<sup>132</sup> In such a case, social unity would be only “apparent, as its stability is contingent on circumstances remaining such as not to upset the fortunate convergence of

---

<sup>131</sup> Rawls, *Political Liberalism*, 9.

<sup>132</sup> Rawls, *Political Liberalism*, 147.

interests”.<sup>133</sup> An overlapping consensus among the different groups in a society would be more likely to survive changes in their comparative political power.

This supplies a new question for political philosophy: can a political conception of justice in fact serve as the basis for an overlapping consensus? One upshot of this new framework is that justice as fairness occupies a slightly less central position in *Political Liberalism* than it did in *A Theory of Justice*. Rawls simply sees justice as fairness as one particularly worthy political conception of justice, and one that may turn out to be particularly appealing and stable in a society with fundamental political ideas like the United States. It is certainly an interpretation of “fair conditions under which the representatives of free and equal citizens are to specify the terms of social cooperation”.<sup>134</sup> Changing course from *A Theory of Justice*, where Rawls appeared to present justice as fairness as a comprehensive doctrine, one that should be accepted because it is true, he now takes pains to present justice as fairness as a freestanding political conception of justice. However, there are societies governed by political conceptions of justice other than justice as fairness. Rawls emphasizes that different sets of ideas could form an overlapping consensus in different societies, depending on the comprehensive views of their members. Certain conceptions of divine right, or even of “dictatorship”, might count as political conceptions of justice, and even serve as an overlapping consensus, though Rawls grants that “the corresponding regimes would lack the historical, religious, and philosophical justifications with which we are acquainted”.<sup>135</sup> However, even granting the possibility of unconventional or even illiberal overlapping consensuses, there is still no guarantee that a society will be able to find an overlapping consensus at all. There may be too much fragmentation among the reasonable comprehensive doctrines to find such a consensus.

---

<sup>133</sup> Rawls, *Political Liberalism*, 147.

<sup>134</sup> Rawls, *Political Liberalism*, 25.

<sup>135</sup> Rawls, *Political Liberalism*, 374 .

## 7 *Political Liberalism*: Political Authority

Curiously, even though it is sometimes said that *A Theory of Justice* is a book about substantive justice, while *Political Liberalism* is a book about legitimacy, *Political Liberalism* says much less about political authority and political obligation than does *A Theory of Justice*.<sup>136</sup> As I mentioned earlier, *Political Liberalism* contains the Liberal Principle of Legitimacy, but this concerns the “exercise of political power”, and he writes elsewhere that “political power is always coercive power”.<sup>137</sup> Thus it does not seem to concern obligation directly. Moreover, he identifies legitimate procedures as those procedures that “all may reasonably accept [them] as free and equal when collective decisions must be made and agreement is normally lacking.”<sup>138</sup> But this does not entail that those procedures are authoritatively binding, or that people in a society must treat them as authoritatively binding. We must do a bit of work, then, to reconstruct the implications of *Political Liberalism* for political authority (at least in Wolff’s sense) and political obligation. When do people have an obligation to obey the law just because it is the law, and how does that obligation relate to their own individual comprehensive doctrines?

First of all, since the Liberal Principle of Legitimacy does not relate to obligation directly, it seems that questions of political obligation are instead going to be left to the *content* of the political consensus of justice and the comprehensive doctrines themselves, rather than to the Liberal Principle of Legitimacy itself. In other words, do the reasonable comprehensive doctrines contain principles of political authority and political obligation? Then, to understand whether Rawls thinks there is political authority in a society with an overlapping consensus, we must revisit the relationship between the values found in an overlapping consensus and in the comprehensive doctrines themselves. In order for there to be an overlapping consensus on some political conception, Rawls suggests that all reasonable comprehensive doctrines actually present

---

<sup>136</sup> Langvaten, “Legitimate, but unjust; just, but illegitimate: Rawls on political legitimacy,” 138.

<sup>137</sup> Rawls, *Political Liberalism*, 68.

<sup>138</sup> Rawls, *Political Liberalism*, 428.

in a society must “endorse”<sup>139</sup> or “affirm”<sup>140</sup> that political conception.<sup>141</sup> This suggests that they must judge the content of the political conception not only to be permissible but also required. For the most part, this would imply that values in the individual comprehensive doctrines will never contradict the political consensus of justice.<sup>142</sup>

On this first level, then, *Political Liberalism* lacks a theory of political authority at all. This is because Rawls never requires people to ignore their own comprehensive views in favour of the political conception of justice. To some extent, this makes the task of finding an overlapping consensus even harder. Without authority to override someone’s comprehensive view, Rawls is forced to construct the overlapping consensus very narrowly to avoid these conflicts. This is why it is so crucial to remove the most divisive questions from the domain of the political conception of justice, and perhaps gives us reason to doubt whether it is at all possible to construct such an overlapping consensus.

While Rawls does not present any reasons why principles of political authority *must* be found within an overlapping consensus, he clearly believes they are *consistent* with an overlapping consensus, so that an overlapping consensus could in theory contain them. Indeed, this appears most clearly in Rawls’s discussion of a “constitutional consensus”, which he presents as an alternative for societies that are unable to reach consensus on substantive values.<sup>143</sup> He describes a constitutional consensus as a “constitution satisfying certain basic principles [that] establishes democratic electoral procedures for moderating political rivalry

---

<sup>139</sup> Rawls, *Political Liberalism*, 134.

<sup>140</sup> Rawls, *Political Liberalism*, 145.

<sup>141</sup> In other places, Rawls suggests the political values in the overlapping consensus are “not viewed as incompatible with basic religious, philosophical and moral values.” (Rawls, *Political Liberalism*, 157)

<sup>142</sup> However, there is one remaining way that they might contradict one another, because Rawls grants that certain comprehensive doctrines allow for internal value conflicts. For example, Rawls discusses the possibility of a “pluralist” view that leaves its “values to be balanced against one another, either in groups or singly, in particular kinds of cases.” (Rawls, *Political Liberalism*, 145) Yet Rawls does not believe that such cases pose a problem for a political consensus of justice. He writes, “its political values normally outweigh whatever other values oppose them, at least under the reasonably favorable conditions that make a constitutional democracy possible.” (Rawls, *Political Liberalism*, 155)

<sup>143</sup> Rawls, *Political Liberalism*, 158.

within society.”<sup>144</sup> (That said, he does little to defend why so many comprehensive doctrines would come to accept a content-independent intrusion into their substantive commitments.) More generally, there is no inherent reason why a principle of authority, a principle requiring someone to exclude certain reasons from their deliberation about how to act, could not be one of the various principles found in a political conception of justice.

## 8 *Political Liberalism: No Solution to Wolff’s Challenge*

Because questions of political authority and political obligation are resolved internally to the comprehensive doctrines in *Political Liberalism*, we still need an account of how each comprehensive doctrine could endorse a principle of political authority that requires those who endorse it to sometimes exclude certain moral considerations from their deliberation. As we have already seen from the failure of moralist views to provide a satisfactory theory of political authority, this account has proved illusive. The situation is even worse when we imagine a society actually attempting to reach a political consensus on its principles of political obligation. While it is plausible that a society with different comprehensive doctrines might be able to reach a consensus on certain substantive values, it is much harder to see how it could reach a consensus on a principle of authority, one that requires them to exclude certain considerations from deliberation and instead treat the law itself as a source of reasons. Thus, *Political Liberalism* suffers from the same problems as moralism, just multiplied many times across many comprehensive doctrines.

I will conclude with one general point about the account of political authority in *Political Liberalism*. As I noted at the beginning of my discussion of Political Liberalism, perhaps its most important feature is that it does not assume a well-ordered society. It instead presupposes a pluralistic society in which people endorse a wide range of reasonable comprehensive doctrines. This leads to a kind of inversion between *A Theory of Justice* and *Political Liberalism*. In *A Theory of Justice*, Rawls sets out to find a moral theory that determines what people should or

---

<sup>144</sup> Rawls, *Political Liberalism*, 158.

should not do, but one that must pass certain tests of stability. In contrast, in *Political Liberalism*, Rawls sets out to show the possibility of a stable social order, but one that must pass certain moral tests. As a result of this inversion, two books are answering fundamentally different questions. But this inversion ultimately dooms *Political Liberalism* as an explanation for how to reconcile authority and autonomy. The only sense of authority in *Political Liberalism* is *de facto* authority (though perhaps with some moral constraints on it, like Rawls's insistence that the relevant consensus be drawn from *reasonable* comprehensive doctrines). Rawls is no longer talking about what people should in fact do, but what people believe that they should do, though subject to certain tests of reasonability. For example, in any society that is partially formed of anarchists, there will be no political authority in that overlapping consensus.<sup>145</sup> Thus the argument of *Political Liberalism* leaves us further away from a liberal reconciliation of political authority and autonomy.

## 9 Conclusion

Despite Rawls's acknowledgement of the value of autonomy, it seems that *A Theory of Justice* and *Political Liberalism* do not adequately show that political authority can be reconciled with it, at least as long as we are working with Wolff's demanding definition of legitimate authority. I argued that the view in *A Theory of Justice* fails to explain why people should exclude contrary moral considerations in deciding what to do. The view in *Political Liberalism*, in contrast, fails to be a theory of political authority at all. It fails to explain why people should actually obey authorities at all if they do not believe that they should do so.

One might wonder: even though Rawls's two-step version of moralism fails, might another two-step version of moralism succeed? It is hard to see how this could be the case, however. I do not have a general argument against two-step versions of moralism, but to make the general claim more plausible, it will help to consider another version of moralism. Suppose that instead of using the original position to pick out their moral principles, an agent simply

---

<sup>145</sup> I think there is no obstacle to anarchism in Wolff's sense being a reasonable comprehensive doctrine, since it does not seek to impose itself on other reasonable doctrines.

chooses their moral principles and takes their choice (rather than the features of the principle) as *wholly constitutive* of what makes the principle binding. Suppose then that one of the principles that the agent picks is a principle requiring them to not act on certain considerations that speak against what the law demands. It seems clear that Wolff would not regard such a way of choosing principles as acceptable; this would ultimately collapse into a form of consent theory.

In order for the liberal to show that the state possesses legitimate authority, then, they will need to look beyond moralism and find some other kind of justification for why someone must exclude moral considerations from deliberation – a standpoint from which it can make sense to evaluate and reject the importance of acting on those moral considerations. This might sound somewhat airy, an abstract point about metaethics rather than political theory, but in the next chapters, I will argue that Raz and Darwall both seek exactly such a standpoint in their theories of political authority. Raz, seeing the obstacles to a completely moral justification of political authority, instead attempts to ground political authority in structural principles of rationality. Darwall attempts to ground authority in the “second-person standpoint”. Despite their ingenuity in reaching beyond moralism, however, I will argue that these attempts to escape moralism are not successful.

## Chapter 5

# The Service Conception of Authority

### 1 Introduction

In this chapter, I present Joseph Raz's service conception of authority as a possible solution to Wolff's challenge. As we have seen in previous chapters, the moralists (including Rawls) lack an explanation of why someone could permissibly exclude morally relevant considerations while acting.

Raz's solution to this problem begins with the ingenious observation that exclusion is not merely found in authority. There are many examples of exclusion in ordinary reasoning. For example, an investor will need to take into account the fact that they are exhausted in deciding whether to take a risky investment. For Raz, similar considerations can explain why authority has force. This is encapsulated in the best-known part of his view, the normal justification thesis, which says that subjects should regard directives as authoritative (and exclusionary) when conforming to those directives would help them better conform to their background reasons than if they relied on their own judgment. For example, Raz would say that insofar as drivers have reason to get where they are going safely and quickly, they must accept the authority of a traffic light – sometimes even when it is mistaken. Thus the reason that we have to obey authorities comes from the pressure to conform to right reason in general.

I begin by presenting Raz's concept of an exclusionary reason, which is the key component of his theory of authority. An exclusionary reason, in short, is a reason to disregard or not act upon some other reason for action. Next, I present his view of authority, which has two parts. First, he presents a descriptive, conceptual analysis of what authority is and what authoritative directives are intended to do, an analysis partly in terms of exclusionary reasons. Second, he describes the normative aspects of authority in a set of several theses: when authority is legitimate (the normal justification thesis and the independence condition), what factors authoritative directives should be based upon (the dependence thesis), and how authoritative directives should be taken by those subject to them (the pre-emption thesis). These theses explain why people might be entitled to treat the directives of authorities as exclusionary reasons.

I then argue that the service conception of authority seems to fare better than moralism in addressing Wolff's challenge. Unlike moralists, Raz has a plausible explanation for why citizens can not act on some of the factors that seem morally relevant in a situation. In short, if obedience would help them conform to right reason, they would be irrational not to do so. Thus it appears, at least at first glance, that the service conception of authority has reconciled authority and autonomy.

## 2 Exclusionary Reasons

Raz's service conception of authority rests on some of his earlier work concerning the foundations of practical reasoning, so I will need to say something about the latter in order to explain the former. Raz began to develop his ideas about authority and practical reasoning at a period in which philosophers were just beginning to use the methods of conceptual analysis to analyze the relationships between the concepts involved in practical reasoning, like rationality, reasons and normativity. At that time, philosophers had reached a fairly wide consensus that questions about what one ought to do should be understood in terms of reasons: we ought to perform an action when the balance of reasons favours it.

In *Practical Reasons and Norms*, Raz argues that this picture is incomplete without a distinction between first-order and second-order reasons. When people talk about reasons for action, they usually have in mind a "first-order" reason – some quality of that action that speaks directly in favour or against doing it. For example, the fact that an action promotes the general happiness, or the fact that an action would advance some of the agent's ends, are both first-order reasons for performing that action.

He argues that not all reasons for action take this form. Consider someone who is looking to invest some of their money.<sup>146</sup> Suppose that this person must make a decision about whether to take a promising investment opportunity within a few hours, but they are presently exhausted and unable to make an effective decision. How should she take the fact that she is exhausted into

---

<sup>146</sup> Raz, *Practical Reason and Norms*, 37.

consideration in her reasoning? It seems like it must play some role in her reasoning, but it is not properly speaking a property of the investment, and so not a first-order reason. Instead, this fact is a “second-order” reason, that is, a reason about a reason. There are many kinds of second-order reasons, but the investor’s fatigue is a particular kind of second-order reason called an “exclusionary” reason. An exclusionary reason is a reason to “disregard” our other reasons.<sup>147</sup> He clarifies in the postscript to the second edition of *Practical Reason and Norms* that he means that exclusionary reasons are reasons “for not acting” on certain reasons, or reasons “for not being motivated in one’s action by certain (valid) considerations”.<sup>148</sup> In the case of the investment, for example, the fatigue gives the would-be investor a reason to not act on their reasons in favour of making the investment.<sup>149</sup> He thinks that the existence of exclusionary reasons shows that there are some cases in which it is rational for an agent to act contrary to the balance of reasons, to exclude certain facts in deciding how to act.

### 3 The Service Conception of Authority: Conceptual Aspects

Having made the case in *Practical Reason and Norms* that exclusionary reasons are a normal part of our reasoning, and that agents are entitled to rely upon them, Raz next argues in *The Authority of Law* that authoritative directives are intended to serve as exclusionary reasons for action. This is the central feature of his conceptual analysis of authority, which understands

---

<sup>147</sup> Raz, *The Authority of Law*, 18.

<sup>148</sup> Raz, *Practical Reason and Norms*, 184-5.

<sup>149</sup> The reader may have some doubts about whether this example is a genuine example of an exclusionary reason. Raz himself grants that his analysis of this example is “incomplete”. It does not necessarily make clear why she rejects rather than accepts the offer. He suggests that in this particular case, the exclusionary reason works like a “rule of thumb,” favouring the rejection of risk over the acceptance of potential benefit. (Raz, *Practical Reason and Norms*, 38) Should the example not work, another example by David Enoch is perhaps a more promising of exclusionary reasons ‘in the wild’. Enoch points out that to the extent that thinking about the calories of a dessert would prevent us from enjoying the dessert, we have reason not to think about that reason at all, though this would be more of an example of a per-emptory reason than an exclusionary reason. (Enoch, “Authority and Reason-Giving,” 321)

authority in terms of a “normative power”.<sup>150</sup> On this view, legitimate authority is the power to change a person’s pre-emptive reasons for action by making authoritative directives, where a pre-emptive reason is a combination of an exclusionary reason and a first-order reason.<sup>151</sup> In other words, when someone exercises legitimate authority over another person, they give that person a new first-order reason to do some action, and at the same time, give them reasons to not act on some of the reasons that spoke against doing that action.

I have already explained the general idea of an exclusionary reason in the previous section, but it is important to be as exact as possible about Raz’s view of the kind of exclusionary reason involved in particular in authoritative directives, since the details are so closely linked to how he addresses Wolff’s challenge to authority. That is, we must be clear about the following question: in what sense are the exclusionary reasons that are given by authoritative directives reasons “not to act” on other reasons? What exactly is it we have reason to not do, think or be?

First, in general, authoritative directives do not prohibit the “reflection on the merits” of the actions they require or proscribe.<sup>152</sup> In other words, even if someone has exercised authority over someone else, the latter person remains free to form their own view about what they should do – as we saw earlier, what matters is their action or motivation for action. Second, the presence of an exclusionary reason does not imply that the excluded reasons no longer have any force at all. Going back to the investment example, Raz would say that the investor’s fatigue does not make it the case that the considerations that speak in favour of the investment no longer speak in favour of the investment. In later work, Raz in fact suggests there might still be reason to conform to those reasons, just not to act on them. Raz suggests that “the best course is if they are indirectly obeyed, i.e. if the action they indicate is performed for some other, independent

---

<sup>150</sup> Raz, *The Authority of Law*, 18-9

<sup>151</sup> Raz actually calls them “protected” reasons, but following his later usage as well as common usage, I will call them “pre-emptive” reasons here.

<sup>152</sup> Raz, *The Morality of Freedom*, 39. Raz notes that authoritative directives could in principle prohibit such reflection, like how military orders are intended to engender immediate compliance. However, Raz thinks that this is not a default or usual feature of authoritative directives, which in turn is reflected in the common expressions that express those authoritative directives. One would actually have to order, “Do five push-ups, and don’t think about it!”

reason”.<sup>153</sup> Third, in cases where a person has multiple sufficient reasons for action, Raz suggests that a person in the grip of an authoritative directive is not under a duty to make the directive their actual motivating reason for action. Someone who has been commanded by another person may have other sufficient reasons, in addition to the command, for doing what is commanded of them.<sup>154</sup> For example, the law tells us not to kill others (in virtually all circumstances), but we presumably all already have independent moral reason to refrain from killing other people. In such a case, it would even be *better* that we act for our own reasons, since a person to be morally suspect if their reason for avoiding murder was that the state demanded so.<sup>155</sup> This seems to mean that the exclusionary reasons created by authoritative directives exclude an agent’s reasons against performing as indicated by a directive, but do not exclude an agent’s independent reasons for performing as indicated by the directive.<sup>156</sup>

Relatedly, Raz suggests that the fact that authorities give people pre-emptive reasons helps to explain how authoritative directives could affect a subject’s reasons even when the subject already had sufficient reason to perform the directed action. If authoritative directives simply gave people new first-order reasons for action, it would be difficult to see what they could add to a situation in which someone already has sufficient reason to do as directed. But authoritative directives can add something to a situation in which someone has sufficient reason to do something in virtue of “their ability to turn “oughts” into duties.”<sup>157</sup> While an “ought” can simply reflect the sum of reasons for and against different actions, Raz says that one has a duty to perform an action “only if it is required by a protected reason which does not derive merely from the fact that adherence to it facilitates realization of the agent’s goals.”<sup>158</sup> It is the protected or preemptive reason, which includes an exclusionary reason, that is crucial in explaining the

---

<sup>153</sup> Raz, *Practical Reason and Norms*, 185.

<sup>154</sup> Raz, *The Morality of Freedom*, 39.

<sup>155</sup> Raz, *Between Authority and Interpretation*, 144.

<sup>156</sup> Raz, *Between Authority and Interpretation*, 144.

<sup>157</sup> Raz, *The Morality of Freedom*, 60.

<sup>158</sup> Raz, *The Authority of Law*, 235.

novel contribution of authorities in cases like these. Authorities can give subjects reasons not to act on reasons that speak contrary to the action in question. Even if giving someone a protected reason does not necessarily affect what they ultimately should do, it will always affect how they should arrive at the conclusion to do it.

His analysis of orders invokes many of the same ideas that were used to explain the idea of authority. An order is a speech act that is “given with the intention that their addressees shall take [it] as a protected reason for action”.<sup>159</sup> While orders must intend to exclude, at the very minimum, the recipient’s “current desires”<sup>160</sup>, Raz suggests that there is quite a lot of variation between different kinds of orders in terms of what kinds of considerations they exclude. For example, he suggests that many orders issued by parents to their children exclude only a child’s present desires, perhaps since the point of those orders is to remedy the child’s inability to overcome the immediate force of those desires in favour of their slightly longer-term interests.<sup>161</sup> He also suggests in slightly later work that authoritative directives must be “presented” as “someone’s view of how its subjects ought to behave.”<sup>162</sup>

Note that normative powers need not always be exercised by means of orders. As long as there is “sufficient reason” for regarding someone’s actions (whether orders, other speech acts or other actions altogether) as protected reasons to do some action, and “if the reason for doing so regarding it is that it is desirable to enable people to change protected reasons by such acts, if they wish to do so,” then those actions will count as exercises of a normative power.<sup>163</sup> It even follows from this claim that normative powers can be exercised unintentionally.

---

<sup>159</sup> Raz, *The Authority of Law*, 22.

<sup>160</sup> Raz, *The Authority of Law*, 23.

<sup>161</sup> It bears repeating that not all orders are in fact legitimately authoritative. One needs to possess legitimate authority in order to be “entitled” to give such orders (Raz, *The Authority of Law*, 15) The conditions for legitimate authority are given by the normal justification thesis.

<sup>162</sup> Raz, “Authority, Law, Morality,” 303.

<sup>163</sup> Raz, *The Authority of Law*, 18. One might doubt that this clause excludes anything at all from counting as authoritative directive. But it in fact does. For example, it excludes certain cases of power from counting as authoritative. Even if I have the power to put someone under an obligation by blackmailing them, and it was

## 4 The Service Conception of Authority: Normative Aspects

The power analysis of authority is an account of the normative impact that authoritative directives are intended to have on their targets, but it does not say anything about whether or why anyone actually possesses such a normative power. Raz begins to take up the search for these conditions at the end of the first chapter of *The Authority of Law*. There, he considers the famous example of whether someone must stop at a well-lit traffic light in the middle of the night when there is clearly no traffic approaching the intersection. The example is typically deployed to cast doubt on whether the state can ever have legitimate authority, or whether there is a duty to obey the law, since many people have the strong intuition that there is no duty to stop at the stop light if no harm would come to anyone by doing so. But Raz suggests that those who do not think there is a duty to obey the traffic light have failed to note the importance of a crucial aspect of the example. In short, the very reason why it is useful to treat traffic lights as exclusionary reasons is that doing so saves us the time, difficulty and risk of making judgments about traffic for ourselves.<sup>164</sup> But the only way that we can obtain this benefit is if we treat the traffic light as an exclusionary reason for action in all or at least almost all cases, since if we were to form a judgment about whether to stop or go each time we decide whether to obey the traffic light, and leave open the possibility of acting on that judgment, we would not be able to obtain the benefit we hoped to obtain from our deference to the traffic light.<sup>165</sup>

In *The Morality of Freedom*, which Raz refers to as his “most complete discussion” of authority<sup>166</sup>, Raz expands his brief observations about the traffic light example into a fully-

---

generally good that they be placed under that obligation, that would still not count as an exercise of authority, because it is not in general desirable that I have the power to impose obligations on someone by means of blackmail.

<sup>164</sup> Raz, *The Authority of Law*, 25.

<sup>165</sup> One puzzling aspect of the traffic light case is that part of the benefit of accepting the traffic light’s authority is that it eliminates the need to deliberate about whether to cross the street, while in the case of political authority, the subject is still fully permitted to deliberate.

<sup>166</sup> Raz, “On Respect, Authority, and Neutrality: A Response,” 298.

developed account of “the nature and role of legitimate authority” which he calls the “service conception of authority”.<sup>167</sup> Roughly, people can be bound to authorities because those authorities provide them with a service, and that service is to help them better conform to what reason requires of them. Before I get into the full details of how authorities perform this service, it is important to avoid a misconception that people sometimes make about the service conception. Raz’s use of the word “service” might give the impression that submission to authority is optional or chosen, based upon whether one cares for the service it provides, but he does not mean this at all. Even though the point of authority is to provide a service for those subject to it, it is not a service that can be declined by those subject to it, at least when they stand in need of that service. In particular, whether or not they consent to authority plays little or no role on whether they must regard an authority’s directives as binding.

The service conception of authority takes the form of four theses.<sup>168</sup> Perhaps the most well-known and important of the four theses is the normal justification thesis, which describes when someone has (legitimate) authority over another person:

The normal and primary way to establish that a person should be acknowledged to have authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, than if he tries to follow the reasons which apply to him directly.<sup>169</sup>

In short, we should acknowledge that someone (or something) has legitimate authority over us if doing so would help us better conform to our background reasons. For example, the agent in the earlier traffic light scenario has a number of background reasons that apply to him. He wants to get where he is going safely, as quickly as possible, with minimal risk to others, making

---

<sup>167</sup> Raz, *The Morality of Freedom*, 55. Raz offers a lengthy explanation of the kind and point of conceptual analysis involved here in *Between Authority and Interpretation*, 129-34.

<sup>168</sup> Originally it took the form of three (the normal justification thesis, dependence thesis, and the pre-emption thesis), but Raz seems to grant in later work that the independence condition counts as a fourth thesis.

<sup>169</sup> Raz, “Authority, Law, and Morality,” 299. (See also: Raz, *The Morality of Freedom*, 53)

concessions to the needs of others to get where they are going quickly. In such a situation, he would conform better to his reasons if he was to treat the traffic light as authoritatively binding than if he was to attempt to determine for himself whether to proceed through the intersection and act on the basis of that determination.

Note that the normal justification thesis implies that authority will be piecemeal – the state will exert different degrees of authority over different citizens. We can see this by considering what Raz lists as the five “most common reasons capable of establishing the legitimacy of an authority”:

- (1) When the authority is “wiser”.
- (2) When it “has a steadier will less likely to be tainted by bias, weakness, or impetuosity.”
- (3) When direct individual action is likely to be “self-defeating.”
- (4) When the process of deciding would be cognitively costly.
- (5) When the authority can “better achieve what the individual has reason to do.”<sup>170</sup>

Each of these conditions provides a rough heuristic for whether the normal justification thesis applies to a given situation. Raz would say, for example, that when someone has greater knowledge, this makes it more likely that complying with their orders will help one better conform to one’s own reasons. Framing the service conception in terms of these intermediate conditions makes it all the more clear that the normal justification thesis only grants legitimate authority to states in a piecemeal fashion. For example, some governments frequently possess a great deal of expertise, like public health experts, transportation safety experts, and so on, but even that expertise has clear limits. Raz suggests that governments may lack authority with respect to some issues or persons, like an “expert pharmacologist” whom the government may not have authority over with respect to “matters of the safety of drugs”, and “an inhabitant of a little village by a river” who may not be subject to authority “in matters of navigation and

---

<sup>170</sup> Raz, *The Morality of Freedom*, 75.

conservation of the river”.<sup>171</sup> This has the effect of making Raz close to an anarchist on many matters, including moral matters.<sup>172</sup>

While the conditions of the normal justification thesis come quite close to being sufficient for establishing the legitimacy of an authority, Raz hints that some further conditions would need to be met to show that someone’s directives are authoritatively binding. First, he also suggests that it must be the case that no other authority is in a better position to help the subject conform to their reasons.<sup>173</sup> Second, even as early as *The Morality of Freedom*, he writes that the reach of the normal justification thesis is limited by the “intrinsic desirability of people conducting their own life by their own lights”.<sup>174</sup> In later work, he elevates this thought into a thesis that is intended to supplement the normal justification thesis, which he calls the “independence condition”. The independence condition is “premised on the thought that it is important that people decide for themselves how to conduct their lives, and that, especially in some areas, they should do so with only limited reliance on direct advice, let alone commands, from others.”<sup>175</sup> On certain matters that are core to people’s lives, like the comprehensive conception of the good they pursue, who they marry, who their friends are, and so on, it is crucial

---

<sup>171</sup> Raz, *The Morality of Freedom*, 74.

<sup>172</sup> One way of extending the reach of the service conception’s authority beyond mere technical matters is to posit that some people, or some deliberative bodies, possess greater moral expertise. That said, even if some people actually are moral experts, people whose judgments about morality are more likely to be correct than the average person, one must show that their expertise would be available to the subject to make use of it. Suppose that some form of Kantian ethics was the correct account of morality, so that everyone would generally conform better to their moral reasons if they were to conform to the Kantian expert’s commands. A utilitarian would still look at this Kantian authority and fail to realize that the person is a moral expert. No doubt the utilitarian would be making a mistake by failing to obey them, but from their perspective, they wouldn’t be making a mistake. So though an expert’s decisions might be correct by the lights of the correct moral theory, this does not show that such a person’s authority would be acceptable to people who hold incorrect moral theories. This proposal would need some method of identifying individual moral experts that is independent of any moral theory at all.

<sup>173</sup> Raz writes of cases with multiple putative authorities, “In such cases the question whether a given authority’s power extends to exclude the authority of another is to be judged in the way we judge the legitimacy of its power on any matter, namely whether we would conform better to reason by trying to follow its directives than if we do not.” (Raz, *Between Authority and Interpretation*, 143)

<sup>174</sup> Raz, *The Morality of Freedom*, 57.

<sup>175</sup> Raz, *Between Authority and Interpretation*, 149.

that individuals independently decide for themselves. But on other matters, like how to conduct oneself in public, the duty has less weight.

Raz grants that we will not always be able to accurately determine whether a putative authority is in a position to provide us the service of helping us better conform to our reasons. Even if someone has enough technical expertise that following their directives would help us conform to our reasons, they might not have the credentials that we typically use as a heuristic for determining who has that technical expertise. (They might be an auto-didact, for example.) As a result, in certain cases, the benefits of their authority will be inaccessible to us, and Raz grants that this implies that they will not in fact have authority over us. In some cases, like the authority possessed by the state, this problem tends not to arise – we have a long history with state authorities, and we can usually assume that they have access to expertise and will enjoy the compliance of others. In other cases, it inevitably takes “inquiry” to determine who has authority, and the costs of these inquiry must be taken into account in determining whether one has reason to submit oneself to their judgments.<sup>176</sup> One could imagine a borderline case where a putative expert is issuing orders that would give a subject a slight advantage in complying with their reasons, but that the process of checking their credentials is sufficiently strenuous to tip the balance against their being reason to obey their directives.

The third thesis in the service conception is the dependence thesis:

All authoritative directives should be based, among other factors, on reasons which apply to the subjects of those directives and which bear on the circumstances covered by the directives. Such reasons I shall call dependent reasons.<sup>177</sup>

The dependence thesis says that authoritative directives should be based on the underlying reasons of subjects rather than other considerations.<sup>178</sup> It is important to clarify the dependence

---

<sup>176</sup> Raz, *Between Authority and Interpretation*, 147.

<sup>177</sup> Raz, “Authority, Law, and Morality,” 299.

<sup>178</sup> Although Raz frames the dependence thesis as a claim about what authorities *should* do, which might be taken to suggest that it is possible, if unwise or self-defeating, for authorities to do otherwise, I think it is intended to state something close to a necessary condition for a subject to have sufficient reason to act as a putative authority demands. In other words, in order for there to be sufficient reason for me to treat someone’s directive as imposing a

thesis in three respects. First, the dependence thesis states that authorities must base their directives on the reasons that apply to subjects, rather than their interests. It is of course true that a person's interests frequently count as reasons for them to act in a particular way, but it is not automatically true that all of a subject's interests count as reasons for them. Likewise, it seems plausible that subjects frequently have reasons which are unrelated to their interests. For example, given a person's duty to contribute to the military defense of their country, a Razian authority might command them to serve in the army, even if doing so goes directly against their own interest. Second, while legitimate directives must be "based" on the reasons of their targets, the dependence thesis is not saying that the subject must initially have reason to perform the very action that is directed by the authority. For example, he suggests that we might initially lack any distinguishing reason between buying a loaf of sliced bread and a loaf of unsliced bread.<sup>179</sup> Were an authority to choose among these options, and direct the subject to buy the loaf of sliced bread, they would be giving the subject a new reason for action (e.g. to buy a loaf of sliced bread) that nevertheless depends in the appropriate way on the subject's previous reasons (to buy bread in general). He suggests that a similar argument might be made about how authority can contribute to solving coordination schemes and other game-theoretical sorts of problems where there are multiple adequate solutions.<sup>180</sup> In later work, Raz describes this process as "concretization", where we have a general reason that serves as a "background" reason.<sup>181</sup> A group of drivers, for example, all have a general reason to drive in the same direction, but in the absence of an authoritative directive, they do not yet have reason to drive in any particular direction.<sup>182</sup> Third, Raz acknowledges that the authority sometimes must take into account reasons that do not apply to the subject; they frequently need to consider "bureaucratic" factors

---

duty on me, that directive must reflect the reasons that would already independently apply to me, whether I am aware of those reasons or not.

<sup>179</sup> Raz, *The Morality of Freedom*, 48.

<sup>180</sup> Raz, *The Morality of Freedom*, 50-1.

<sup>181</sup> Raz, *Between Authority and Interpretation*, 152.

<sup>182</sup> Raz notes that a putative legitimate authority might need to be a *de facto* authority in order to carry this function out. (Raz, *Between Authority and Interpretation*, 158)

alongside “substantive considerations”.<sup>183</sup> He gives the example of a “de minimis” rule, which excludes courts from ruling on matters that are insignificant, but it might equally be illustrated by the state’s need to issue general “blanket” laws that apply to all persons at all times, even though directives tailored to individual circumstances might have better reflected the reasons of individual subjects.<sup>184</sup>

The final thesis in the service conception of authority is the pre-emption thesis:

The fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them.<sup>185</sup>

In other words, if an authority’s directives depend on the subject’s reasons, as is required by the dependence thesis, the subject can no longer take those original reasons into account in deciding what to do. That is, the subject must not act directly on their background reasons, and instead must take the directives as their reasons for action.

Since the preemption thesis describes authoritative directives as pre-emptive reasons, it seems to cover much the same ground as the analysis of authority as a normative power that he defended earlier in *The Authority of Law*, which states that authority is a power to change someone’s protected reasons for action. Raz himself does not explicitly describe the relationship between the pre-emption thesis and his earlier analysis, but it seems to me that the power analysis complements the pre-emption thesis without contradicting it.<sup>186</sup> First of all, the pre-

---

<sup>183</sup> Raz, *The Morality of Freedom*, 51.

<sup>184</sup> Of course, the state also has rule of law reasons to issue general rules rather than particular directives, and perhaps those are really the primary reason it does so.

<sup>185</sup> Raz, *The Morality of Freedom*, 46.

<sup>186</sup> Another possibility is that he changed his mind between *The Authority of Law* and *The Morality of Freedom*. The earlier account of authority in *The Authority of Law* seems to separate the question of what authority is from how authority is justified or legitimate. In later work, Raz discusses the role of “neutrality” in conceptual analysis of authority, which would require that “the account of authority should explain what follows when someone has authority, but will not include anything about the conditions under which one may acquire or hold authority.” (Raz, *Between Authority and Interpretation*, 132) Raz suggests that such neutrality is undesirable, suggesting that “the way to justify a claim that one has authority” must be guided by “what has to be justified, namely the consequences

emption thesis is framed as a claim about how subjects should respond to directives that are legitimate in virtue of the dependence thesis and the normal justification thesis. The power analysis provides an explanation of what authority is: it is a power to change someone's protected reasons. But if authority has such a power, then it makes sense that a subject should respond to its directives by treating them as pre-emptive reasons. Second, Raz presents different arguments for each of them. While he argued for that authority should be analyzed as a normative power using a direct appeal to our intuitions about the exclusionary nature of authoritative directives, Raz argues for the pre-emption thesis by showing that it follows from the dependence thesis. Raz presents a "double-counting" argument that the pre-emption thesis follows from the dependence thesis. In short, if the directives themselves served as reasons, but did not replace or exclude the underlying reasons that the directive was based upon, the agent would in effect be counting the same consideration twice.<sup>187</sup>

One might think that the pre-emption thesis would only apply when a directive is successfully based on the reasons that it is intended to pre-empt, but Raz does not think that this is the case. Even when directives fail to adequately reflect the subject's underlying reasons, the subject should still allow the directive to replace their underlying reasons, as long as the normal justification thesis as a whole is met. This might happen, for example, if an authority happened to be wrong on a particular occasion but had a long history of being correct. He suggests that mistaken directives must retain their exclusionary force because otherwise they would always be "open to challenge", which in turn would imply that "the advantage gained by accepting the authority as a more reliable and successful guide to right reason would disappear."<sup>188</sup>

The possibility that authoritative directives might be binding on a subject even though they fail to reflect those reasons obviously carries substantial risks. Indeed, Raz grants that this means that sometimes governments will be able to pass "immoral and unjust laws" that are

---

of having authority." (Raz, *Between Authority and Interpretation*, 132). Perhaps even by *The Morality of Freedom* he had come to reject this requirement. Even so, I'm not sure that this has large substantive implications for his view of authority.

<sup>187</sup> Raz, *The Morality of Freedom*, 58.

<sup>188</sup> Raz, *The Morality of Freedom*, 61.

binding on subjects.<sup>189</sup> In these situations, a government can “act wrongly in the exercise of their authority without forfeiting it”.<sup>190</sup> Raz seems to pin his hopes that this will result in relatively little harm on the subject’s continued “sensitiv[ity] to the presence of non-excluded considerations.”<sup>191</sup> Recall that authoritative directives exclude some reasons for action but do not necessarily exclude all of them. If it becomes clear, for example, that the traffic light is malfunctioning, we are entitled to ignore it, because when we choose to do so, we are appealing to reasons other than those that the traffic light is intended to exclude. Likewise, disputants must respect an arbitrator’s judgment insofar as it is intended to sum up the merits of each side in their dispute, but if they discover that the arbitrator was drunk when they made the decision, that gives them reason to disregard the arbitrator’s judgment.<sup>192</sup> So he appears to hope that some of the moral risk of obedience is diminished by the likelihood that unjust laws will tend to be excludable for analogous reasons. And if that does not do enough to prevent subjects from being bound to evil directives, he provides two more arguments that subjects might not be bound to unjust directives. Under certain rare conditions a subject of a Razian authority can disregard that authority’s directives because they judge that the content of those directives is mistaken. First, since subjects of an authority are allowed to rely on an authority’s ‘track record’ in determining who is an authority, they can certainly update their assessment of that ‘track record’ in light of a bad directive, which will affect whether they should treat the authority’s subsequent directives as authoritatively binding. Second, he suggests that certain directives can be so inconsistent with reason that they do not bind subjects even at the time they are given. To this end, he draws a distinction between a “great mistake” and a “clear mistake”.<sup>193</sup> He uses an example of authoritative advice to make this point. If a friend is normally much better at arithmetic and it makes sense to defer to their judgment in all ordinary cases, even when the error appears to have

---

<sup>189</sup> Raz, *The Morality of Freedom*, 79.

<sup>190</sup> Raz, *The Morality of Freedom*, 79.

<sup>191</sup> Raz, *The Authority of Law*, 24.

<sup>192</sup> Raz, *The Morality of Freedom*, 42.

<sup>193</sup> Raz, *The Morality of Freedom*, 62.

a great magnitude, one might still be in a position to judge that the friend has made an error if they suggest that the sum of two integers is a partial fraction.<sup>194</sup>

Altogether, then, the service conception of authority suggests that the role of authority is to help those subject to it to better conform to their background reasons. Both the nature of authority (given by its analysis in terms of a normative power, as well as the pre-emption thesis) and its justificatory conditions (given by the normal justification thesis and the dependence thesis) help it to perform this role. If an authority does not live up to the normal justification thesis, it is (at best) merely *de facto* authority – it claims to be an authority, it may be believed to be an authority, but it does not possess legitimate authority over those whom it commands.

## 5 A Solution to Wolff's Challenge?

Raz appears to have conceived the service conception of authority as an almost explicit response to Wolff's problem, one designed from the ground-up to show how an autonomous person could treat authoritative commands as exclusionary, content-independent reasons. He endorses a principle of autonomy that "entails action on one's own judgment on all moral questions" and seeks a picture of authority that is consistent with that.<sup>195</sup>

At the same time, he seeks to avoid the pitfalls of moralist theories of authority. He doubts that there is any moral argument for political obligation – even for a *prima facie* duty to obey the law – which suggests that he would deny that there is a purely moral argument for political authority either.<sup>196</sup> Indeed, his view of authority differs substantially from moralism. Our duty to obey political authorities, to the extent that it exists, does not follow from one moral duty or another moral duty. Instead, it comes from its capacity to help us conform to reason in general. On the service conception, there is no single correct set of moral values that gives rise to authority.

---

<sup>194</sup> Raz, *The Morality of Freedom*, 62.

<sup>195</sup> Raz, *The Authority of Law*, 3.

<sup>196</sup> Raz, *The Authority of Law*, 237-242.

Though an authority will need to be knowledgeable about the moral reasons that apply to a subject, to differing degrees depending on the matters that its commands concern, in principle the service conception itself is neutral, for example, as to whether utilitarianism or some other moral theory is the correct one. Raz would say that both utilitarians and Kantians would have to endorse the normal justification thesis.

One might think that because an authority will usually have to get a subject's moral reasons correct in order for their commands to be binding, this makes the service conception a kind of moralism. However, a moralist would understand submission to authority as a moral requirement, one that may not be in "service" of the subject at all. In contrast, there is still some sense in which the normal justification thesis is in service of a subject – it helps them act on their reasons, whatever those reasons might be. Put another way, unlike the moralist, who simply thinks that one of our moral duties is to obey an authority, Raz inserts a middle step. Raz would say that we already have certain independent reasons or moral duties, and obedience to authority helps us conform to those moral duties.<sup>197</sup> This difference turns out to be crucial in explaining why moralists are unable to account for exclusion, while the service conception is.

I will close this chapter by arguing that the service conception makes remarkable strides towards reconciling authority and autonomy, beginning with the first and most important problem: how can an agent fulfill their duty to deliberate yet not act on factors that they deem morally relevant? Raz's answer appeals to his idea of an exclusionary reason – a reason not to act on certain reasons. As Raz notes, an exclusionary reason is a genuine reason, and it can make sense for an autonomous agent to act on one without violating their duty to deliberate. He argues that Wolff failed to appreciate this:

[Wolff] tacitly and correctly assumes that reason never justifies abandoning one's right and duty to act on one's judgment of what ought to be done, all things considered. I shall call this the principle of autonomy. He also tacitly and wrongly assumes that this is identical with the false principle that there are no valid exclusionary reasons, that is, that

---

<sup>197</sup> Note that if there were a moral duty to obey an authority, the normal justification thesis would be somewhat trivially satisfied with respect to that authority. That is, obeying the authority would almost automatically help a subject act on the reasons or duties that apply to them, if one of those reasons or duties was to obey the authority.

one is never justified in not doing what ought to be done on the balance of first-order reasons.<sup>198</sup>

This is only a partial answer to Wolff's problem, because if exclusionary reasons exist, we still need to explain why it makes sense for an agent to treat a command as an exclusionary reason, when at first glance it does not appear to be a reason for anything at all. The service conception explains this by saying that this treatment is required as a matter of practical reason, of means-ends rationality. It is permissible to treat a command as a reason to do what is commanded as long as doing so better helps us conform to right reason in general. Raz writes, "[The reasons created by authoritative commands] are there by reasoning analogous (some would say identical) to that which establishes the existence of instrumental reasons: you have reason to do A, doing B (walking to the station, obeying the authority) will facilitate doing A, therefore you have reason to do B."<sup>199</sup> As he notes, "In postulating that authorities are legitimate only if their directives enable their subjects to better conform to reason, we see authority for what it is: not a denial of people's capacity for rational action, but simply one device, one method, through the use of which people can achieve the goal (telos) of their capacity for rational action, albeit not through its direct use."<sup>200</sup> As we have seen, this is likely to occur when someone has superior knowledge, is more neutral, or can achieve coordination that an individual cannot on their own, among other situations. The reasons that help us act on our other reasons really do exist, and it would be a mistake for us not to act upon them. This explanation applies to both moral reasons and non-moral reasons.<sup>201</sup>

I will conclude by raising three doubts that one might have regarding whether the service conception addresses Wolff's challenge. First, one might doubt that the commands given by

---

<sup>198</sup> Raz, *The Authority of Law*, 27.

<sup>199</sup> Raz, "On Respect, Authority and Neutrality," 299. Note that I have changed the letters in the quote to X and Y to avoid confusion.

<sup>200</sup> Raz, *Between Authority and Interpretation*, 140.

<sup>201</sup> Of course, this may not be fully satisfactory to Wolff, who rejected reliance on expertise as much as authority. However, if Raz is correct, the ball is now back in Wolff's court: if an agent really does have exclusionary reasons under the normal justification thesis, it is the anarchist who is deliberating badly, not the Razian.

authorities under the service conception are genuinely content-independent because their force ultimately hinges on their relationship to right reason. However, this relationship is indirect, and Raz thinks that when the normal justification thesis is satisfied, an authority's commands are binding even when they are wrong. Thus, I think that the service conception of authority satisfactorily accounts for the content-independence of commands, at least in the sense understood by Hart.

Second, one might doubt that the service conception is content-independent because it does not apply to every action that an authority might command. For example, it wouldn't extend to matters about someone's private life. This is true, but this doubt relies upon an implausible understanding of content-independence – any reasonable theory of political authority would grant that there are some spheres of human life, however narrow, that are off limits to it. The drill sergeant's command is not any the less content-independent even if she doesn't intend for her command that her subordinate give her all his money to be taken as a peremptory or exclusionary reason. Nor is the force of the traffic light any less content-independent because it can only shine red, yellow or green.

One might also doubt whether the service conception accurately deals with the *moral* problems associated with obedience. Does it resolve the fact, for example, that one person will be subject to the will of another person? To some degree, I think Raz attempts to address this problem with his independence condition, a constraint on the normal justification thesis which says that “it is important that people decide for themselves how to conduct their lives, and that, especially in some areas, they should do so with only limited reliance on direct advice, let alone commands, from others.”<sup>202</sup> However, this reflects a genuine disagreement with Wolff. For Wolff, there may well be moral issues with turning over any of one's choices to another person, whereas for Raz, these issues are content-dependent, hinging on which choices are being turned over. However, I argued in chapter 2 that this is not Wolff's chief complaint about obedience. The main obstacle to reconciling authority and autonomy is to explain how a subject could make sense of excluding moral considerations in their assessment of what to do, and how a subject could treat commands as reasons even though they appear to be content-independent.

---

<sup>202</sup> Raz, *Between Authority and Interpretation*, 149.

## 6 Conclusion

In explaining the central features of the service conception of authority, I hope to have shown why many philosophers have been attracted to it. At least at first glance, the service conception of authority can make sense of why an agent could neglect certain morally relevant factors in deciding what to do without violating their duties of autonomy. It does so by finding a standpoint outside of morality which can displace moral considerations – the structural, content-independent standards of practical rationality in general. In the next chapter, drawing on work by Stephen Darwall, I will argue that this is not a genuine solution to Wolff's problem after all. The service conception adequately explains exclusion and content-independence, but does not actually describe authority, least of all the kind of authority that is possessed by the state.

## Chapter 6 Darwall's Objection to the Service Conception

### 1 Introduction

In this chapter, I present Stephen Darwall's objection to the service conception of authority. His objection to the service conception relies on two main counterexamples. The first counterexample is an expertise case, one in which the normal justification thesis appears to be satisfied because the subject lacks the expertise that would enable them to conform to right reason on their own. If John is an experienced cook, and Sara wants to cook the best possible meal, it seems the normal justification thesis would say that John has authority over her, since she would better conform to her reasons by treating John's instructions as exclusionary reasons that displace her own judgment on cooking. Yet Darwall argues quite persuasively that it is implausible that John has authority over Sara, since Sara remains free to disregard John's instructions as she sees fit. Thus, it serves as a counterexample to the normal justification thesis. In reply, Raz argues that the normal justification thesis does not apply to the case of John and Sara after all, and so it is not actually a counterexample to the normal justification thesis. I argue that his reply is not successful.

The second counterexample is a motivation case, one in which the normal justification thesis is satisfied because the subject lacks the motivation to do as reason demands them to do. For example, if I need to get out of bed in the morning, but cannot convince myself to do so, I might try to hire a person to issue me an order to wake up. Once again, it is implausible that the person whom I have hired has any genuine authority over me. I might do well to treat them as though they have authority over me, but they do not actually have authority over me. Darwall's discussion of the second counterexample gives us a clearer idea of what he thinks goes wrong with the service conception of authority. For Darwall, the mark of legitimate authority is that it is truly "illegitimate" for the obedient person to consider factors aside from the command itself in their deliberations.<sup>203</sup> This gives us some insight as to why the service conception is mistaken.

---

<sup>203</sup> Darwall, "Authority and Reasons: Second-Personal and Exclusionary," 273.

Even if it would be irrational for me to disobey an authoritative command, as would be the case when the conditions of the normal justification thesis are satisfied, I remain free in some sense to disregard the authority's directives. In contrast, if someone commands me, and their command has the force it is intended to have, then I am not free to disregard their directives.

I then discuss the best possible case for the service conception of authority: the case in which obeying an authority can help a subject act on their *moral* reasons. In this case, it plausibly does seem illegitimate for the agent not to treat the commands of the authority as exclusionary reasons. However, even in this best case scenario, I argue that the normal justification thesis is not why it is illegitimate for the subject to decide and act for themselves.

Finally, I introduce Darwall's view of what *would* be required to give rise to authority-based exclusionary reasons: some sense in which an authority has a "right" to rule. I say more about his view of authority in the next chapter, but in this chapter I argue that Raz's attempt to argue that the normal justification thesis gives rise to a right to rule is unsuccessful.

Overall, we can see from Darwall's objections that Raz's attempt to ground legitimate authority in practical reason is not successful. While Raz was correct to think that practical reason contains some form of exclusionary reasons, genuine authority involves a different kind of exclusion. It is a kind of exclusion that does not have as its target conformity with right reason. Thus, practical reason does not provide the standpoint that explains why an autonomous agent can obey an authority and exclude their competing reasons for action.

## 2 The Normal Justification Thesis

It will be useful to quickly revisit the most important part of the service conception of authority: the normal justification thesis. The normal justification thesis says that someone has legitimate authority over another person if the latter person would better conform to reason by treating the former's directives as authoritatively binding than by trying to reason for themselves.<sup>204</sup> The

---

<sup>204</sup> Raz, "Authority, Law, and Morality," 299. (See also: Raz, *The Morality of Freedom*, 53)

normal justification thesis attempts to ground authority in practical reason. We should treat authoritative directives as binding because it is rational to do so. I argued that the great attraction of this view is that it offers an apparent explanation for authoritative exclusion, like the way we are expected to disregard other features of a situation and just follow a traffic light, that is distinct from the purely moralist explanation. If doing so can help us better act on our reasons in general, it seems reasonable to let the traffic light play this role in our reasoning.

One way of arguing against the normal justification thesis is to present a counterexample to it. This would take the form of a case in which it claims that someone would have legitimate authority, but it is clear that the person does not have legitimate authority. Darwall's argument against the service conception relies on two such counterexamples. While Darwall was not the first philosopher to raise a counterexample against the normal justification thesis, his explanation of his counterexamples is perhaps the most carefully developed, and is also the one that drew the most developed response from Raz.<sup>205</sup>

### 3 The Expertise Counterexample

The first counterexample was originally raised by Raz himself as a possible objection to the service conception of authority, but it was one to which he thought he had an adequate response. Imagine that John is an expert chef, and that his friend Sara would like to cook a meal (and has reason to do so).<sup>206</sup> Seeing that Sara needs his help, John begins issuing directives about how to best cook a meal, like to properly slice the fish along its length. According to the normal justification thesis, John possesses legitimate authority over Sara, since she would better comply with the reasons that already apply to her if she treated his directives as pre-emptive reasons, which implies that he has authority over her. Nonetheless, it seems intuitively clear that John does not have any kind of practical authority over Sara. There's some crucial sense in which it is

---

<sup>205</sup> For example, Scott Shapiro raises a similar example. (Shapiro, "Authority," 416) Also, as Scott Hershovitz notes, Donald Regan used a similar counterexample in an even earlier piece. See Regan, "Authority and Value: Reflections on Raz's Morality of Freedom".

<sup>206</sup> Raz, *The Morality of Freedom*, 64. Note that I follow Darwall in assigning the names John and Sara to the case.

still “up to her” whether to conform to his directives, so she doesn’t seem to be in the grip of anything like practical authority.

Raz acknowledges that the case of John and Sara appears at first glance to be a counterexample to the normal justification thesis. However, he thought that he could show that John and Sara do not actually meet the conditions of the normal justification thesis. He writes that, “But whether or not there is a complete justification for me to regard their advice or instructions as guides to my conduct in the way I regard a binding authoritative directive depends on my other goals.”<sup>207</sup> In short, he seems to be saying that John’s directives are not “authoritatively binding” on Sara because Sara could get out of them simply by changing her goals. For example, Raz suggests that “if I wish to enjoy myself dabbling in cooking.... then I should try to form my own judgment [instead]”.<sup>208</sup> This would imply that John and Sara do not really satisfy the conditions of the normal justification thesis at all, and so their case is not a genuine counterexample to the normal justification thesis.

Darwall argues that Raz’s response to his own counterexample is inadequate. To slightly elaborate on his reasoning, the question of whether Sara can *change* her goals is strictly speaking irrelevant to whether they satisfy the conditions of the normal justification thesis at a particular point in time. At the point in time at which Sara has “no reason to do anything other than prepare the best meal”, the normal justification thesis would apply to her.<sup>209</sup> Later, after Sara changes her goals, so that her new goal is to experiment with cooking, Raz is correct that the normal justification thesis would no longer apply to her. But that does not stop the normal justification thesis from applying to her at the first point in time, and so the normal justification thesis still problematically implies that John must have authority over her in that first point in time.<sup>210</sup> Thus

---

<sup>207</sup> Raz, *The Morality of Freedom*, 64.

<sup>208</sup> Raz, *The Morality of Freedom*, 64.

<sup>209</sup> Darwall, “Authority and Second-Personal Reasons for Acting,” in *Morality, Authority and Law: Essays in Second-Personal Ethics*, 147. Darwall puts this point in terms of the ability to “stipulate” Sara’s goals for the sake of argument.

<sup>210</sup> More generally, the normal justification thesis assigns authority at an instantaneous point in time – it does not tell us anything about what it is to hold authority *over time*.

Raz has not successfully showed that John and Sara do not meet the conditions of the normal justification thesis, and so their case remains a counterexample against it.

In reply to Darwall, Raz gives a new argument that John and Sara do not satisfy the normal justification thesis. He observes that once John has issued his directives, assuming that Sara believes that John is an expert in cooking, John and Sara “no longer meet the condition of the NJT”.<sup>211</sup> This is because after John has issued his directives, it is no longer true that John knows “what she should do better than [Sara] does”.<sup>212</sup> Raz’s argument, then, seems to be the following:

- (1) Once John directs Sara, she knows as much as John about what to do in her situation.
- (2) If she knows as much about John about what to do in that situation, Sara would not do better by treating John’s directives as pre-emptive reasons than she would by attempting to ascertain what she should do for herself.
- (3) Once John directs Sara, she would not do better by treating John’s directives as pre-emptive reasons than she would by attempting to ascertain what she should do for herself.
- (4) Therefore, once John directs Sara, they no longer satisfy the conditions of the normal justification thesis.

In short, because Sara knows as much as John once John has directed her, she no longer needs to treat his directive itself as a pre-emptive reason for action – he has effectively stepped out of the picture. Raz then draws a more general conclusion from the case of John and Sara: expertise in general is never enough to satisfy the conditions of the normal justification thesis. Instead,

---

<sup>211</sup> Raz, “On Respect, Authority and Neutrality,” 301.

<sup>212</sup> Raz, “On Respect, Authority and Neutrality,” 301.

expertise must be “mixed with other considerations, such as the need for coordination, for concretizing indeterminate boundaries, and the like” in order to ground legitimate authority.<sup>213</sup>

However, Raz’s reply to Darwall is not successful. This is for two reasons. First, it is not clear that premise (2) of his argument is true. The problem is that even though there is a sense in which Sara knows what to do just as much as John, after he has given his directive, Sara would still seem to be treating John’s directive as a pre-emptive reason for action. She still does not have access to the underlying reasons that explain why John’s directive is correct. Compare the situation from John’s perspective and Sara’s perspective. If John were to cut the filet lengthwise himself, his reasons for doing so would be facts about *fish*, like “the bones tend to break if you cut it width-wise”. In contrast, Sara’s reason for action is about *John*. It’s simply “John told me to do it”. Moreover, since “John told me to do it” is a reason that sums up other reasons, Raz must hold that she cannot treat John’s directive simply as a first-order reason to be weighed alongside the other first-order reasons that she sees. She must treat it as a pre-emptive substitute for her own first-order reasons, otherwise she will be guilty of double-counting some reasons (e.g. her own suspicions about the bones in the filet).

Second, as Scott Hershovitz notes, even if the premises were true and the argument were valid, the argument would extend to most cases of legitimate authority, and authority would risk becoming vanishingly rare.<sup>214</sup> Raz would almost become an anarchist, which is surely not his intent. Raz intends for his argument to only apply to expertise cases, but his argument would also extend to coordination cases. To see why, consider the classic example of a coordination case: the way that a state might determine which side of the street motorists should drive on. Raz would suggest that this is precisely the kind of case where we might better conform to our background reasons (e.g. to get where we want to go quickly and safely) by treating the law as pre-emptive. However, given what Raz has already said about expertise cases, it is hard to see how he could treat coordination cases any differently from expertise cases. In the same way that John would have ‘stepped out of the picture’ once he had issued his directives to Sara, it seems

---

<sup>213</sup> Raz, “On Respect, Authority, and Neutrality,” 301.

<sup>214</sup> Hershovitz, “The Role of Authority,” 8.

that the authority in a coordination case likewise ‘steps out of the picture’ once it has issued its commands. Thus if Raz is going to say that John does not have legitimate authority over Sara, he must also say that the state does not have legitimate authority over its citizens in determining which side of the road to drive on.

That said, because Hershovitz is responding to an earlier draft article by Raz, one where Raz suggests that coordination cases are the sole remaining case of authority, he does not notice that there are some cases of the normal justification thesis that escape this second argument.<sup>215</sup> These are motivation cases, cases where someone lacks motivation to do what they already have reason to do.<sup>216</sup> In these cases, it seems like the authority can’t just ‘step out of the picture’, because the subject is counting on the authority’s will (or haranguing) to get them to do what they already have reason to do. I think this is why when Darwall makes his central argument against Raz, he switches his examples, using a motivation case rather than his original expertise case.

## 4 The Motivation Counterexample

Darwall’s second counterexample concerns a case where someone lacks motivation to do what they already have reason to do. Consider B, who has very strong prudential reason to get out of bed in the morning (like not wanting to miss class), but also has some competing reasons to stay in bed (like its warmth and coziness). Although B overall has much stronger reason to get out of bed, he finds every morning that he is just unable to get himself to act on this stronger reason. Finally, at wit’s end, he hires A to wake him up and order him out of bed, knowing that he will be motivated to follow A’s demands. According to the normal justification thesis, it appears that A would actually have authority over B. This is because B would better conform to his reasons

---

<sup>215</sup> Hershovitz is responding to Raz, “The Possibility of Partiality”. This article was never published, though its arguments were taken up in a number of published articles.

<sup>216</sup> Here I draw on Raz’s original classification of cases of the normal justification thesis (Raz, *The Morality of Freedom*, 75), as well as Scott Shapiro’s commentary on it. (Shapiro, “Authority,” 405-6)

by treating A's orders as pre-emptive reasons. Yet it seems intuitively clear that the A does not possess any legitimate authority over B.<sup>217</sup>

While this second counterexample might not seem importantly different from the first one, Darwall uses it to make a further, rather different point about the service conception. He argues that not only does A lack legitimate authority, but also A's commands do not even count as genuinely exclusionary reasons for B. He denies that meeting the conditions of the normal justification thesis guarantees that an exclusionary reason is created:

...And perhaps he has even weightier pragmatic reasons not even to think about the reasons in favour of continuing to lie in bed. But however weighty those pragmatic reasons might be, they do not constitute or create preemptive reasons or authority-based exclusionary reasons. Even if he would be an idiot to think for a moment of the reasons for staying in bed, it is in no way illegitimate or beyond his discretion to do so.<sup>218</sup>

What Darwall seems to be presenting here is a *test* for whether an exclusionary reason really is present: if B actually has an exclusionary reason to do some action, then it should be "illegitimate" or "beyond his discretion" to think of the reasons against performing that action.<sup>219</sup> (To avoid an obvious reply by Raz, I will slightly modify Darwall's test: it must be the case that it is illegitimate to *act* on the reasons against performing the action.<sup>220</sup> A person remains free to think about their reasons; what matters is how they act.) This test suggests that B does not have any exclusionary reasons – it is not "beyond the discretion" of B to act on the reasons keeping him in bed. While it might be unwise for him to do so, he is not forbidden from doing so in any sense.

---

<sup>217</sup> Darwall, "Authority and Reasons: Second-Personal and Exclusionary," 273.

<sup>218</sup> Darwall, "Authority and Reasons: Second-Personal and Exclusionary," 273.

<sup>219</sup> Darwall, "Authority and Reasons: Second-Personal and Exclusionary," 273.

<sup>220</sup> As I note, Darwall's actual test for the presence of an exclusionary reason is that it should be "illegitimate" or beyond someone's discretion for them to think of the reasons against taking the action recommended by the directive. Raz points out that this test misunderstands the notion of an exclusionary reason, which is a reason to not *act* on a reason, rather than a reason to not *deliberate* on it. (Raz, "On Respect, Authority and Neutrality," 298n21) But this doesn't seem to be of any further consequence for their disagreement.

In his reply to Darwall, Raz doubles down on his claim that B really does have an exclusionary reason. B should stop trying to decide for themselves whether to get out of bed, and just rely on A, if doing so would help them get out of bed.<sup>221</sup>

Between Raz and Darwall, who is correct about whether B has an exclusionary reason? On the one hand, I think that Raz is correct that there is *some sense* in which B has an exclusionary reason. B really does have reason to disregard his own judgment and instead treat A's orders as reasons for action. As such, I think that Darwall is mistaken when he argues that the conditions of the normal justification thesis only show that someone is better off treating the directives of the authority *as though* they provide exclusionary reasons. The reasons exist because, in general, when one has reason to achieve some end, one has reason to take the means to that end. If B has reason to wake up early in the morning, and would do better at waking up early in the morning by obeying A, B has reason to stop allowing their own assessment of the reasons to play a role in whether they get up, and instead to rely on the judgment of A.<sup>222</sup> Raz's thought seems to be that this principle of means-end rationality applies not only to ordinary reasons, but also to exclusionary reasons, and Darwall does not seem to present a reason to think that an ordinary exclusionary reason is not present in this case.

On the other hand, I think that Darwall is quite correct that this is not the sense of exclusion involved in genuine authority. In fact, he already more or less concedes that his test for the presence of an exclusionary reason is mistaken when he notes that some exclusionary reasons are "authority-based" and others are not. He suggests that it is not necessarily the case that the exclusionary reason to not get out of bed is an "authority-based" exclusionary reason, that is, an exclusionary reason that would "remove the decision from one person to another."<sup>223</sup> Exclusion

---

<sup>221</sup> Raz, "On Respect, Authority and Neutrality," 299.

<sup>222</sup> Raz, "On Respect, Authority and Neutrality," 299. That said, there is one important difference between means-end rationality and the normal justification thesis. The normal justification thesis is not restricted to cases where obedience is a "necessary" means in some end. It is enough that obedience simply be better than trying to work things out for oneself.

<sup>223</sup> Darwall, "Authority and Reasons: Second-Personal and Exclusionary," 272.

in practical rationality is different than the kind of exclusion involved in authority, which we can see in three respects.

First, one important thing to notice about the motivational counterexample is that B would likely have an exclusionary reason against acting on some of his reasons even if A was not present. Anyone who has a habit of reasoning poorly would already have an exclusionary reason forbidding them from acting on some of their reasons as they seem to them. For example, if B was guaranteed to assign too much weight to the reasons in favour of saying in bed, then B already had reason to compensate for this in his reasoning, perhaps by weighing them less heavily than he otherwise would have. This is helpful for seeing that there is nothing distinctively authoritative about B's exclusionary reason.

Second, it is important to remember that an exclusionary reason is just that – a *reason* to not treat certain considerations as reasons. The most that can be said of someone who fails to act on such a reason, on that description of the situation alone, is that they are irrational. In light of this, it is unlikely that a simple exclusionary reason is likely to bind someone's will in the way that genuine authority does.

Third, Raz lacks a principled explanation of why an obedient person should exclude *all* considerations from their deliberation about how to act, and simply act according to the authority's directives. Most of the time, someone who took the command into account in their reasoning, but who also continued to act on the underlying reasons *in favour* of the commanded action, would presumably fare even better in conforming to their reasons.<sup>224</sup> Yet intuitively, the kind of exclusion that is present in authority is exclusion to not act on *all* one's underlying reasons, both those that speak in favour and those that speak against the action. This is the force that the traffic light appears to possess.

This specific point reflects a more powerful criticism that is sometimes made of the service conception, that the service conception of authority confuses 'counsel' and 'command'

---

<sup>224</sup> Being able to act on the reasons that speak in favour of a commanded might not *always* confer this advantage in conforming to their reasons (e.g. in cases where the agent's deficiency is their ability to calculate the reasons), but it seems like it would sometimes, with respect to some agents.

(or advice and command), or that it confuses epistemic (or theoretical) authority and practical authority.<sup>225</sup> Even though the service conception is to some degree content-independent, in that it holds that directives can be binding even when they are mistaken, the force of directives still ultimately depend on their relationship to Sara, in the same way that advice ultimately depends on its relationship to Sara. In general, in taking someone's advice, we often have reason to not act on our own judgment and instead rely on the person whom is offering the advice. If someone has a much better history with the stock market than I do, I should simply take their advice on which stocks to pick, rather than try to choose for myself. It would be irrational for me to do otherwise, even if my advisor unfortunately happens to be wrong on this particular occasion. In contrast, the kind of exclusion involved in genuine practical authority is not in service of conformity with right reason; an obedient person should be prepared for the possibility that what they are commanded to do will contradict their reasons, even if all uncertainty were removed from the situation. It is this relationship with Sara that also explains why Raz lacks a principled explanation for why authorities should prevent a subject from acting on all of their reasons, not just the reasons against the commanded actions. If an authority's goal is simply to bring a subject into better conformance with right reason, then the subject should be free to use all means at their disposal to achieve this goal.

In response, Raz has suggested that it is a mistake to associate the service conception with epistemic authority. Epistemic authority concerns belief, while practical authority concerns action, and the service conception of authority is about action.<sup>226</sup> While this might be true, it does not remove the force of the objection. Even if Raz is correct that the normal justification thesis is about people's reasons for action rather than their reasons for belief, it seems to change their reasons for action *by* changing what reasons for belief would be appropriate for them to hold. Even if it is not necessary for the agent to actually change what they believe, it seems that

---

<sup>225</sup> Hobbes most famously draws this distinction between counsel and command: "it is manifest that law in general is not counsel, but command; nor a command of any man to any man, but only of him whose command is addressed to one formerly obliged to obey him." (Hobbes, *Leviathan*, 173). See Darwall, *The Second-Person Standpoint*, 13n25, 15n29, for more discussion of this objection.

<sup>226</sup> Raz, *Between Authority and Interpretation*, 155.

when the conditions of the normal justification thesis apply, it will be rational for the agent to form the belief that what is commanded is correct.

Altogether, we can see that Raz's attempt to explain the exclusion found in legitimate authority by appealing to the familiar exclusion found in rationality is not successful. The normal justification thesis, as a part of practical reason, gives rise to its own kind of exclusion, which is not the kind of exclusion present in genuine authority.

## 5 Cases with Moral Dependent Reasons

There is one remaining set of cases in which the normal justification thesis might still appear to have binding force. These are cases in which the subject's underlying or dependent reasons are moral. Suppose, for example, that B must get out of bed to meet a critical deadline for filing documentation for a client's asylum claim, but again, has to hire A to get himself out of bed. Now that the stakes in this revised example are moral, doesn't B have a binding obligation to treat A's commands as exclusionary reasons in Darwall's sense? Wouldn't it be "illegitimate" or "beyond his discretion" to do otherwise?

Darwall considers this case, but I don't think that his response is adequate. He grants that it would be illegitimate for B to do otherwise, but he says that A's demand is no different than the "legitimate demand" of a third party. As he writes, "But none of this gives A any distinctive authority to create pre-emptive reasons for B that other members of the moral community do not have."<sup>227</sup> However, it seems clear that there could be lots of reasons why A, and not other members of the moral community, might have this status. For example, it might be the case that B has special respect for A, or that he is terrified of A. For Raz this is purely a matter of B's psychology: what sort of features must A possess for his orders to make B more likely to act on the reasons that apply to him? There is no reason to think that every member of the moral community will possess these factors.

---

<sup>227</sup> Darwall, "Authority and Reasons: Exclusionary and Second-Personal," 277.

I have a different response than Darwall.<sup>228</sup> In the original case, which concerns only non-moral reasons, B only has an exclusionary *reason*, but in this new case, I think that B instead has an exclusionary moral *duty*. An exclusionary moral duty is a moral duty (rather than just a reason) to exclude certain considerations from his reasoning, and to just rely on A's orders instead. Exclusionary moral duties are much stronger than exclusionary reasons; if B has an exclusionary moral duty, this really would imply that it would be "illegitimate" for him to do otherwise. Crucially, exclusionary moral duties have a different source than exclusionary reasons. They are not brought into existence by practical reason; one doesn't have an exclusionary moral duty to help them conform to reason in general. Instead, whatever principle governs exclusionary moral duties must be a moral one: when someone has a moral duty, they must take the necessary means to fulfill it. Unlike the normal justification thesis, this is not a principle of pure practical reason, because it does not apply to prudential and other non-moral reasons. Instead, it is a principle of substantive moral theory, one that can be outweighed or directly contradicted by other moral duties (like a moral duty of autonomy). So contrary to Raz, it is not the original normal justification thesis that is doing the work in bringing about the exclusionary moral duty; instead it is its fully moralized analogue.

However, even granting that it would be "illegitimate" for B to not follow this exclusionary duty, Darwall would still be correct to deny that this exclusionary moral duty is authority-based. Many of the other distinctive marks of legitimate authority are not present. As I mentioned in the last section, authority is often indifferent to whether it helps a subject better conform to the reasons that apply to them. To reframe the point that I made there in terms of morality rather than rationality: obedience will often appear to the subject to be outright immoral, and an obedient person should be prepared to contradict the moral duties that they previously thought to be pressing. This would be true even under conditions of full information, if the subject knew all of the non-moral facts that might be relevant. So overall, perhaps the right way of thinking about this case is that it involves a third, distinct kind of exclusion. It is not the

---

<sup>228</sup> As a separate point, a picture of Razian authority that distinguished sharply between moral and non-moral reasons would be very strange. On such a view, the normal justification thesis somehow "turns on" as soon as the underlying reasons are moral. The subject would not always be in a position to know whether an authority is issuing deciding cases on the basis of their moral reasons or their non-moral reasons, and so would not know whether the authority is potentially legitimate or what attitude they should take towards its directives.

kind of exclusion found in rationality, nor is it the kind of exclusion that is truly distinctive of authority. Instead, it is a kind of exclusion found purely in the bounds of moral theory.

## 6 The Right to Rule

So far, I have followed Darwall in arguing that there is an element of authority that is not captured by the normal justification thesis. According to Darwall, the missing element is a *right* to direct someone's behaviour.<sup>229</sup> Darwall has a rather complex view of these rights which I will discuss in much greater detail in the next chapter. My question in this chapter is much simpler: does the normal justification thesis show that authorities possess a right to rule? If so, perhaps Raz could explain how authorities give exclusionary reasons after all.

In an unpublished but widely circulated piece, Raz argues that the normal justification thesis nevertheless does give rise to a certain kind of right to rule, though a right that is very different from a right in Darwall's sense. Raz suggests that when there is a duty to obey the authority, one that is owed to the authority, the authority has a right to be obeyed.<sup>230</sup> So determining whether there is a right to rule is a matter of determining whether there is a duty to obey them that is owed to them. In general, Raz suggests that duties are owed to a particular person "when their justification turns on the fact that they protect and promote an interest of the person to whom they are owed", and this implies that the person has a right "which is protected by those duties".<sup>231</sup> Raz then suggests that we might explain the right possessed by authorities in

---

<sup>229</sup> As Darwall writes, "But what *right* could John have in such a case? Or recall Raz's remark that, unlike merely justified coercion, genuine practical authority involves "an appeal for compliance" and an "invocation of the duty to obey" (Raz 1986:25-6). How, in such a case, could John warrantably expect that one would have any obligation to follow his instructions, however foolish one might be not to do so? It follows from our stipulations that John's instructions are not mere advise; they provide pre-emptive reasons. But it is hard to see how that gives John any right to [Sara's] compliance with his directives or [Sara] any obligation to comply with them." (Darwall, "Authority and Second-Personal Reasons for Acting," in *Morality, Authority and Law: Essays in Second-Personal Ethics*, 147)

<sup>230</sup> Raz, "The Possibility of Partiality," 17. Here I follow (with some departures) Scott Hershovitz's description of the argument in "The Role of Authority", 8-9.

<sup>231</sup> Raz, "On Respect, Authority, and Neutrality," 391. (The account of rights in this article closely reflects the account in "The Possibility of Partiality", but provides some useful additional details.)

terms of special interests that authorities have in virtue of the role they play as authorities. As the service conception implies, these interests are not the private interests of the person who holds authority over another person. For example, while anyone might have an “interest” in having public funds deposited in their bank account, this is not one of the interests possessed by an authority. Instead, an authority has an interest “to be a good authority”.<sup>232</sup> With this argument, Raz attempts to show that those who meet the conditions of the normal justification thesis have a “right to rule” after all.

At first glance, it might appear difficult to settle whether Darwall or Raz is correct without settling the even harder question about how to define rights in general. In particular, as I mentioned, Darwall’s own conception of rights is embedded in his framework of the “second-person standpoint”. It would be a quite serious undertaking to show that Darwall’s whole framework is correct. Fortunately, I think we can show that Raz is mistaken without making very substantive assumptions about rights. Even on a fairly minimal understanding of a right, Raz has not successfully shown that the conditions of the normal justification thesis give rise to a right to rule.

The basic problem with Raz’s argument that the interests of authorities in being a “good authority” give them a right to rule is that there is a gulf between ordinary interests and an authority’s interest in being a “good authority”. Consider my very ordinary interest in not experiencing pain. This interest plays a direct role both in the argument that other people should not cause me pain and the argument that I have a right not to be caused pain. On the other hand, the interest in being a good authority plays no role at all in determining what the subject should do. It only shows that the duty to obey, *if it exists*, is owed to the authority figure. But it plays no role at all in explaining whether I should obey the authority or not. It is parasitic on some independent argument that I should obey the authority.<sup>233</sup>

---

<sup>232</sup> Raz, “The Possibility of Partiality,” 18.

<sup>233</sup> This is perhaps another way of putting Herskovitz’s observation that Raz’s argument seems to be “question-begging”. (Herskovitz, “The Role of Authority,” 9)

Scott Hershovitz has identified a further problem with Raz's argument. The argument fails to show that the normal justification thesis establishes the existence of a right to rule even given the way that Raz understands the relationship between rights and interests. Raz suggests that rights exist to promote the interests of the person to whom the corresponding duties are owed. But as Raz himself notes, not all interests are important enough, or of the right kind, to give rise to rights. In order to give rise to a right, the interest must be "sufficient to ground a duty to protect or secure that interest in a significant way."<sup>234</sup> To use a famous example, a sick person's very significant interest in having Henry Fonda come wipe their brow does not give them any rights against Henry Fonda.<sup>235</sup> Why think that an authority's interest in authority is going to be important enough to give rise to a right? As we have already seen, the normal justification thesis can apply even when the matters that it concerns are trivial. If I want to get more fit at the gym, that is enough to give me reason to treat my personal trainer's directives as authoritative. In cases like these, where the matters are trivial, it is especially hard to see how my personal trainer's interest in being a good authority would make it the case that he has a *right* to direct me. So Raz does not provide an explanation for why the interests of an authority in being a good authority are important enough to give rise to a right.<sup>236</sup>

## 7 Conclusion

I have argued that Darwall's objections against the service conception of authority are mostly successful. What these objections ultimately show, though Darwall does not present things exactly in these terms, is that authority cannot be grounded purely in the principles of practical reason. If the worst that we can say of someone who disobeys an authority is that they are being irrational, the authority is not a legitimate one, and it is not binding in the way that authorities are

---

<sup>234</sup> Raz, "The Possibility of Partiality," 9.

<sup>235</sup> Thomson, "A Defense of Abortion," 55.

<sup>236</sup> Hershovitz, "The Role of Authority," 9.

supposed to be. Authority involves a different kind of exclusion than the kind of exclusion that is found in the structural principles of rationality.

To clarify, the argument of this chapter does not imply that Hart is mistaken when he identifies content-independence and peremptoriness (or exclusion) as features of how commands are meant to be taken. This is because he does not claim that they are *constitutive* of how they are meant to be taken – there might be something else that a subject needs to do to respond to commands as they are intended. Nor does the argument of this chapter contradict Wolff, who only claims that obedience involves doing what an authority says because they say so. Wolff does not claim that these are the constitutive features of obedience, although he might be taken to say that they are the only *morally impermissible* features of obedience. The objection of this chapter is only a problem for the service conception of authority, and perhaps certain moralist views. For example, the objection might also extend to fair play theory, since the fact that two people would have to obey a third person out of considerations of fairness to each other does not seem to establish that this third person has any authority to direct their behaviour.

Because the service conception of authority fails to be a theory of authority at all, it is not a genuine solution to Wolff's challenge after all. A liberal must search elsewhere for a way of reconciling legitimate authority and autonomy.

## Chapter 7 Second-Personal Authority

### 1 Introduction

In this chapter, I consider Darwall's own "second-personal" view of authority. The sense in which it is 'second-personal' might be put in this way: without the will of other people to whom we are accountable, nothing would be truly binding on us, or at least our obligations would be very limited.<sup>237</sup> He might say, for example, that if someone were the only person in the world, they would have no binding obligations at all, that they would be completely free to do as they like.<sup>238</sup> This insight has important implications for a view of authority. He says that for authority to actually bind its subjects, for it to actually exclude their competing reasons, those subjects must be accountable to the authority. That is what it means for an authority to have the "right" to make demands of its subjects, and fundamentally why the normal justification thesis failed as a theory of authority.

One of the main reasons why Darwall's view of authority is so interesting is that it seems at first glance to offer an alternative to grounding political authority in practical reason or in morality. We have already seen from his criticisms of the service conception that he would not think that second-personal authority can be understood in terms of the requirements of practical reason. There are also good reasons to hope that second-personal authority can be separated from

---

<sup>237</sup> Note that when Darwall says that authority is second-personal, he does not just mean that it involves another person. We can see this because a first-personal account of authority would be perfectly coherent. Such an account would say that people are only ever accountable to themselves, and that their will (rather than the will of others) makes it the case that they must obey other people. Throughout *The Second-Person Standpoint*, Darwall considers first-personal views of moral obligation as an alternative to his own second-personal account of it.

<sup>238</sup> While I find this to be a useful way of stating the overall spirit of his view, I grant that it is unclear that Darwall would want to say that *all* bindingness can be traced back to the second-person standpoint. There might be some considerations that are binding independent of whether they are demanded by anyone, in a relationship of accountability. For example, he suggests that if someone were to step on another person's foot, and cause them pain, they would have a "state-of-the-world" reason to remove their foot, associated with an impersonal reason to reduce the total amount of pain in the world. (Darwall, *The Second-Person Standpoint*, 5-6) But I am not persuaded that this reason is not derived from moral obligations, which Darwall understands in second-personal terms. Either way, not much hinges on putting his view in such a strong way. What matters for the chapter is the weaker claim that the wills of others are *a* source of binding obligations, even if they are not the only one.

morality. Not all interpersonal demands or relationships of accountability are matters of morality, even they do seem to bind those to whom they apply. However, despite this initial hope, I will argue that second-personal authority does ultimately collapse into either a form of moralism or a form of consent theory (which, as we saw in chapter 1, would also be unacceptable to Wolff).

I begin by explaining how Darwall's view of authority is part of a broader effort to shed light on "second-personal" aspects of our lives – cases where we address one another in a way that presupposes a relationship of accountability and respect, including any case in which we have a right to make a demand of another person. As he understands authority, it is a relationship of accountability between two people in which the will of the person or body issuing the demands, or at least whomever it represents, play an ineliminable role in the creation of a reason to do as it says.

Much of Darwall's discussion concerns cases where authority is symmetrical, like where two people can each make the same demand of one another. He says relatively little about the asymmetrical relations of authority like political authority, and much of my effort in this chapter is directed towards reconstructing what he would say about these relations. I argue that there are two ways of extending the second-personal view of authority into a theory of political authority.

The first theory of political authority simply applies the general second-personal definition of authority to the particular case of political authority. Just as one person can be accountable to another person, this first theory would say that a citizen is accountable to their state. However, I argue that this theory collapses into moralism or consent theory. Most immediately, this can be seen from Darwall's chosen examples of authorities, like the case where someone steps on another's foot, or the example of a waiter and customer. However, it is not just his particular choice of examples that links his theory of authority to moralism. He thinks that all second-personal phenomena carry certain moral presuppositions about respect and agency; indeed, the ultimate conclusion of the book that best details his view, *The Second-Person Standpoint*, is that our second-personal involvement with one another presupposes some version of the Categorical Imperative and contractualism. This is also reflected in his suggestion that every "bipolar" obligation (a purely second-personal obligation, owed to one individual) still entails a moral obligation (an obligation that is owed to the whole moral community).

In other work, Darwall instead describes legal authority as a power to “find” or “make” law. This suggests a second theory of political authority, one that implies that political authority involves not just a right to make claims and demands that reflect existing bipolar obligations, but also a power to *create* bipolar obligations that are owed to third parties, as well as moral obligations more generally. However, I suggest that this view collapses into moralism even more straightforwardly than the first view.

Altogether, then, the second-person standpoint does not turn out to be a genuine alternative to moralism because it collapses into moralism or consent theory.

## 2 The Second-Person Standpoint

Stephen Darwall’s theory of authority is part of a much larger project aimed at describing the irreducible “second-personal” aspect of much of our practical reasoning and its relationship to moral obligation. In order to understand his view of authority, it is necessary to understand the outlines of this project, which I will describe over the next two sections of this chapter.

Darwall defines the “second-person standpoint” as the perspective that we “take up when we make and acknowledge claims on one another’s conduct and will”.<sup>239</sup> He illustrates this idea with the example of someone (say, Abigail) who is intentionally standing on the foot of another person (say, Pablo) and causing him pain.<sup>240</sup> In this example, Abigail already has reason not to cause Pablo pain – pain is from any point of view a bad thing, and the world should contain less of it. Suppose now that Pablo *demand*s that Abigail lift her foot off of his foot. According to Darwall, because Pablo has a right to make this demand, his demand gives Abigail a further “second-personal” reason that Abigail she did not have before, one that is supposed to increase the total amount of reason that she has to lift her foot off of Pablo’s foot. The second-personal reason created by Pablo’s demand also differs in character from Abigail’s pre-existing reason not

---

<sup>239</sup> Darwall, *The Second-Person Standpoint*, 3.

<sup>240</sup> Darwall’s original example does not include the names ‘Pablo’ or ‘Abigail’; I have added them for ease of exposition.

to bring more pain into the world. The second-personal reason could *only* have been brought about through an authoritative claim or demand. It is a fundamentally directional reason, one that is owed to Pablo, such that she is accountable to Pablo for conforming to it. At the same time, this relationship of accountability also makes appropriate certain reactive attitudes, like guilt and resentment, should Abigail not conform to Pablo's demand that she remove her foot from his foot.

Darwall suggests that the concepts raised in this example – accountability, a second-personal reason, authority, a right, and a valid claim – form an irreducible “inferential circle”.<sup>241</sup> In other words, each of them can be, and indeed must be, defined in terms of the others. Whenever we attribute a right to someone, or attribute authority to them, or draw any further inferences from these conclusions, we must do so on the basis of other second-personal concepts. Moreover, the attribution of any of these concepts to someone involves making a certain uniform set of assumptions about them. Most significantly, when we attribute a right to someone, or make a claim on another person, we presume that the person is autonomous, since all second-personal concepts make “a claim on the will”, and purport “to direct a person through her own free choice and in a way that recognizes her status as a free and rational agent.”<sup>242</sup> Such claims presuppose that the person whom they address is capable of being moved by reasons, including considerations that are not merely their own desires, and that they are capable of holding themselves responsible for what they have done.<sup>243</sup>

While the second-personal standpoint is an important part of our practical reasoning, not all of our reasons, and not all the ways we might give another person a reason, are second-personal. There are many reasons (like reasons about our own states of mind, or about states of the world) that are not second-personal. And there are some ways of “giving” reasons that do not count as second-personal. First, advice does not make claims upon our will; it simply points to

---

<sup>241</sup> Darwall, *The Second-Person Standpoint*, 12.

<sup>242</sup> Darwall, *The Second-Person Standpoint*, 49 (see also page 20). Note that Darwall denies that “participants in second-personal interaction invariably do accept or are aware of these presuppositions, or even that the necessary assumptions must be accessible to them.” (24)

<sup>243</sup> Darwall, *The Second Person Standpoint*, 21.

reasons for belief that would exist independently of whether their author had advised or provided testimony. Second, coercion simply “purports to create reasons in something like the way that legitimate claims or demands do, that is, second-personally, but without the appropriate normative backing for the threatened “sanctions,” which consequently provide only the superficial appearance of an accountability relation.”<sup>244</sup> That said, Darwall seems to suggest that all reason-giving that is not advice or coercion is second-personal, so the category of second-personal reasoning still includes a wide range of our practical reasoning and interactions with others.<sup>245</sup>

### 3 Some Distinctions within the Second-Personal: Bipolar Obligations and Moral Obligations

Darwall considers two familiar kinds of reasons and argues that they are second-personal. The first are moral obligations. According to Darwall, moral obligations result from special kinds of implicit demands made by *all* agents on one another. Darwall joins a number of other philosophers, including Mill, Brandt, Gibbard, and Scanlon, in taking this to be the distinctive feature of morality.<sup>246</sup> This implies that morality’s overridingness or categoricity, which many other philosophers treat as its distinctive features, are demoted to derivative features of it – features that are explained by the fact that they are owed to others, rather than the other way around. Altogether, Darwall describes this conception of morality as “morality as equal accountability”, according to which “moral norms regulate a community of equal, mutually accountable, free and rational agents as such, and moral obligations are the demands such agents

---

<sup>244</sup> Darwall, *The Second-Person Standpoint*, 51.

<sup>245</sup> One of the places that Darwall describes this threefold categorization is here: “To hold that impersonal reactive attitudes like blame do not presuppose representative authority is to be committed to the dilemma that either blame is purely epistemic, seeking simply to uniform its object of a moral standard or of the legitimate demands of that the others have the authority to make or process, or that it has practically directive force, but only brutally or nakedly [that is, coercively].” (Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics*, 37)

<sup>246</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics*, 21.

have standing to address to one another and with which they are mutually accountable for complying.”<sup>247</sup>

Second, in more recent work, Darwall considers what are sometimes referred to as “bipolar” moral obligations.<sup>248</sup> Bipolar obligations differ from ordinary moral obligations because they necessarily involve two moral agents, so that one cannot speak coherently of a bipolar obligation that only involves one person. In a bipolar obligation, one person (the obligor) has a duty which is owed to another person (the obligee), and a violation of the obligation can be said to “wrong” the obligee. The duty to keep a promise is a paradigmatic example of a bipolar obligation. When someone has made a promise to the obligee, the obligation to keep the promise is owed to the obligee in particular. Since bipolar obligations are “owed” to someone, Darwall thinks that they more or less wear their second-personality on their sleeve. The obligee has a right that the obligor carry out the promise, and is entitled to hold them accountable for doing so.

While both bipolar obligations and moral obligations must be understood in second-personal terms, they bind their holders in different ways. In a moral obligation, it is the will of the moral community, or the representative member of the moral community, that binds the obligated person. This draws upon a “fundamental de jure normativity” which exists independently from any other relationships.<sup>249</sup> In other words, the fact that the moral community demands some action of us itself gives us a second-personal reason of some form to do that action. In contrast, in a bipolar obligation, it is the will of the obligee that binds the obligated person. It does so directly, not simply because the relevant action is morally required. Bipolar obligations are thus “inherently normative” and have “inherent moral force in themselves” in a way that is at least partially independent from their relationship to moral obligation.<sup>250</sup> Darwall illustrates the way in which bipolar obligations have their own normativity with the example of a

---

<sup>247</sup> Darwall, *The Second-Person Standpoint*, 101.

<sup>248</sup> Darwall credits Ernest Weinrib and Michael Thompson for introducing this term, and notes that these duties are sometimes referred to as “relational” or “directed” obligations or duties. (Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 20)

<sup>249</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 19.

<sup>250</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 25.

rule-utilitarian system of promising. He imagines that such a system could contain obligations that appear superficially similar to genuine bipolar obligations: it would have rules that give promisees standing to hold their promisers accountable for failing to follow through with the promise, the standing to release promisers from their obligations, and so on, precisely because these rules tend to maximize happiness over time. Yet Darwall would deny that such a system contains genuine bipolar obligations.<sup>251</sup> In any genuine bipolar obligation, the will of the obligee is doing work in bringing about the binding force of the obligation.

Bipolar and moral obligations differ in other respects as well. First, while both moral obligations and bipolar obligations are second-personal and so involve reactive attitudes, the reactive attitudes that are involved are different. Bipolar obligations involve the personal reactive attitudes, like guilt and resentment, that we have already seen. In contrast, moral obligations involve impersonal reactive attitudes, which are “impersonal or disinterested or generalized analogues” of [the personal reactive attitudes], including moral indignation, disapprobation...” which are “felt from as if from an uninvolved, third party’s standpoint”.<sup>252</sup> Second, bipolar obligations and moral obligations differ in the *way* in which authorities bring them into existence. Bipolar obligations are brought into existence when someone explicitly makes a claim or demand. For example, in the case of Pablo and Abigail, it is Pablo’s demand that Abigail remove her foot from his foot that creates the bipolar obligation, and until he presses that demand, she does not have a bipolar obligation to remove her foot. In contrast, the general moral obligation to remove her foot from Pablo’s foot seems to precede Pablo’s explicit claim or demand. As Darwall writes, “Moral demands do not come into existence through being made in blaming someone.”<sup>253</sup> Instead, Darwall thinks that impersonal blame involves “add[ing] our voice to or second[ing], as it were, a demand that we must presuppose is made of everyone by the moral community or representative persons as such.”<sup>254</sup> Third, while the claims involved in bipolar obligations are discretionary, the claims involved in moral obligations period are

---

<sup>251</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 25.

<sup>252</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 22.

<sup>253</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 23.

<sup>254</sup> Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 37.

mandatory. For example, while Pablo has a choice whether to demand that Abigail remove her foot off of his foot, members of the moral community are always understood as implicitly demanding that Abigail remove her foot off of his foot whether or not they choose to do so.<sup>255</sup>

How are bipolar obligations and moral obligations related to one another? On the one hand, Darwall thinks it is possible (though rare) that there might be a moral obligation without an overlapping bipolar obligation – for example, there might be a moral obligation to protect the Grand Canyon that is not owed to anyone in particular.<sup>256</sup> On the other hand, every bipolar moral obligation entails a corresponding moral obligation. Darwall’s argument for this entailment is complex, but his basic thought is that to be under a bipolar obligation, a person must be capable of blaming themselves, and to be capable of blaming themselves, that person would have to see themselves as blameworthy from the impersonal moral point of view, from the perspective of the moral community as a whole.<sup>257</sup> Darwall seems to think that when a person sees themselves (or should see themselves) as blameworthy from the impersonal point of view, this would make that impersonal blame actually appropriate, so it follows that the person would be under a moral obligation.<sup>258</sup> So when Pablo demands that Abigail remove her foot from his foot, she is placed

---

<sup>255</sup> This includes Pablo himself, insofar as he is acting as a member of the moral community.

<sup>256</sup> As Darwall writes, “So far as the concept of moral obligation (period) is concerned, there might be obligations that are not owed to anyone, or at least, that go beyond any that are. Perhaps there is an obligation not wantonly to destroy beauty or not to foul that environment that is like that.” (Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 20)

<sup>257</sup> Why think that someone under the grip of a bipolar obligation must also be under the grip of a moral obligation period? Darwall provides an argument for this inference, which can be schematized in the following way:

(1) If someone is under a bipolar obligation, their acknowledgment of that obligation must involve “an internal acceptance of the complaint’s legitimacy”. (Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 38)

(2) Internal acceptance of this obligation requires him to be capable of blaming himself, “not just as from his victim’s standpoint, but as from anyone’s”. (Darwall, “Bipolar Obligation,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 38)

(3) The fact that any arbitrary member of the moral community would be entitled to blame him is equivalent to saying that he is under a moral obligation period.

<sup>258</sup> This seems to make Darwall vulnerable to exactly the kind of objection he raises against the normal justification thesis as a theory of when someone has the power to create pre-emptive reasons. There, he suggested that the fact

under a corresponding moral obligation to do so; we should understand the whole moral community as demanding that she comply with Pablo's demand.

## 4 Second-Person Authority

This discussion of the second-person standpoint, and its application to bipolar obligations and moral obligations, provides us with enough background to understand Darwall's own general account of authority, an account that is intended to describe authority across all of its different contexts. We have already seen the rough outline of this account: authority is at stake when one person makes claims or demands on another person, such that the latter person is accountable to the former person for complying with those claims or demands. To be precise, Darwall defines authority in the following way:

Someone has practical authority with respect to another if, and only if, the latter has a second-personal reason to comply with the former's valid claims and demands and is accountable to the former for so doing.<sup>259</sup>

Darwall writes that these the claims made by authorities might be on either the "conduct" or the "will" of another person.<sup>260</sup> This suggests a fairly wide range of possible ways in which an authoritative claim or demand might be intended to influence someone. A person might simply demand that their listener perform a certain action, without caring why they ultimately perform that action. But a person might also make a claim on the formation of a person's will, so that the command itself is the reason rather than some other consideration. For example, it seems that in the case of Pablo and Abigail, it matters that Abigail lifts her foot off of Pablo's foot *because he said so*; if she did it for another reason, like needing to head to her next foot-stepping

---

that someone would do better by treating consideration as a pre-emptive reason does not imply that that consideration really is a pre-emptive reason. Analogously, here, a person might need to see themselves as blameworthy from the impersonal point of view, but it doesn't follow that they actually are blameworthy from the impersonal point of view.

<sup>259</sup> Darwall, "Authority and Reasons: Exclusionary and Second-Personal," 266.

<sup>260</sup> Darwall, *The Second-Person Standpoint*, 3.

engagement, it would fail to fully acknowledge the force of his directive and express disrespect towards him. Regardless of whether a directive requires that its listener act or form their will in a certain way, the recipient of an authoritative claim is called upon to act because the person exercising authority said so.

We have also already seen that authority in this sense is conceptually tied to accountability. The duty is owed to the former person, in the sense that they will be accountable to them for following through with it, and we will say that that former person has the “right” to direct the conduct of the other person. Accountability in turn is bound up with the reactive attitudes, like guilt and resentment, as well as with respect. Darwall elaborates, “if the private fails to heed the sergeant’s orders, he doesn’t simply act contrary to a reason that sheds favourable light; he violates the order and so disrespects the sergeant and her authority.”<sup>261</sup> This accountability is what explains the normativity of the second-personal reasons created by demands, that is, how and why those demands are binding on us. It is because second-personal reasons “bear conceptually on what we can be legitimately held responsible for doing” that they bind us.<sup>262</sup> Darwall’s thought, then, is that authority partakes in exactly the same kind of binding force because it is interpersonal and involves one person making a claim on another person’s will. Darwall also suggests that part of the sense in which second-personal reasons are binding is that they are at least sometimes pre-emptive.<sup>263</sup> He writes, “Although I agree with Raz that the capacity to create preemptive reasons (that is, to create exclusionary reasons not to act on reasons that would otherwise be unimpeached along with a new reason that “displaces” or pre-empts the excluded reasons) is a mark of practical authority, I believe that this capacity itself requires the second-personal relation of accountability.”<sup>264</sup> So any time that someone makes a claim on

---

<sup>261</sup> Darwall, *The Second-Person Standpoint*, 60.

<sup>262</sup> Darwall, “Law and the Second-Person Standpoint,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 173.

<sup>263</sup> While Darwall thinks that second-personal reasons are pre-emptive, it is not clear whether he thinks this exhausts their direct effect on our reasoning. He might say, for example, that the second-personal aspect of second-personal reasons itself plays some direct role in how they bind people independent from its relationship to pre-emption.

<sup>264</sup> Darwall, “Authority and Reasons: Exclusionary and Second-Personal,” 261.

another person, that person is expected to allow the claim to pre-empt in some sense their own judgment about how to act.

How might second-personal authority be justified? That is, what makes it the case that the relevant relationship of accountability is actually in place when someone makes a demand on another person? Darwall's basic insight on this question is that since authority is second-personal, the conditions that give rise to authority must themselves be second-personal. As he writes, paraphrasing Bernard Williams, "Second-personal authority out, second-personal authority in."<sup>265</sup> In other words, to show that one person has authority over another person, we must show to begin with that they stand in some other morally significant relationship to one another, one that involves a kind of address. This provides another explanation for why the normal justification thesis was mistaken (although one closely linked to the one in the previous chapter). Two people might fulfill the conditions of the normal justification thesis without addressing one another in any meaningful way. I might better conform to my reasons by treating someone's commands as pre-emptive reasons even if those commands are not addressed to me, for example if I were to overhear them by accident. (Of course, it is implausible that they are fully exclusionary. Requests, for example, are second-personal, but presumably requests do not exclude every reason for action.)

We can say more about Darwall's views on the justification of authority by considering his distinction between *representative* authority and *individual* authority. Representative authority is the authority at stake in moral obligations in general, while individual authority is the authority at stake in bipolar obligations.

Representative authority is possessed by all moral agents, where every moral agent is implicitly understood as addressing claims to every other moral agent to conform to their moral obligations. The underlying relationship of accountability in representative authority is simply the basic relationship of accountability that holds between all moral agents. As we have seen, the exercise of representative authority is not up to the discretion of moral agents to make claims or

---

<sup>265</sup> Darwall, "Authority and Second-Personal Reasons for Acting," in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 142.

demands on other agents. For example, to the extent that there is a standing moral obligation not to stand on Pablo's foot, it is understood as the product of the moral community exercising their representative authority through implicit demands on one another.<sup>266</sup> Because representative authority is implicitly claimed, questions of legitimacy do not really come up regarding it, but we can assume that a claim of representative authority would be legitimate when it is addressed to the demands that actually do concern the whole moral community.

In contrast, individual authority is associated with bipolar obligations. It is exercised at the discretion of the person who holds it. For example, Pablo exercises a claim of individual authority on Abigail when he demands that she remove her foot from his foot. It is his choice as to whether or not he should make this demand. In earlier work, Darwall suggests a necessary condition that must be satisfied for a claim of individual authority to be successful: the relevant background relationship of accountability must be "accepted" by both parties in order for the claim to be successful.<sup>267</sup> In the case of a teacher administering a pop quiz, for example, Darwall suggests that what makes that threat legitimate is that "all agree" that the threat is a permissible one in the context of the teacher-student relationship.<sup>268</sup> Likewise, in discussing the example of a sergeant demanding that a private perform push-ups, Darwall emphasizes the "free" character of the private's acceptance of the authority of the sergeant as crucial for why her demand is justified.<sup>269</sup> More recently, Darwall puts the condition in terms of "reasonable acceptance", writing that "Unlike reasons for acting of other kinds, [the existence of second-personal reasons] is staked on its being the case that those to whom they apply can reasonably be expected to

---

<sup>266</sup> Darwall presumably thinks that there is a discretionary, explicit form of representative authority, too. For example, if I were a third party to the situation involving Pablo and Abigail, I could explicitly make a claim on Abigail ("Hey, get off of his foot!") that invokes my representative authority as any member of the moral community. (That said, Darwall would probably say that this representative claim does not give its addressee any *new* reason for action.)

<sup>267</sup> Darwall, *The Second Person Standpoint*, 51.

<sup>268</sup> Darwall, *The Second Person Standpoint*, 43.

<sup>269</sup> Darwall, "Law and the Second-Person Standpoint," in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 175.

accept their validity by accepting the authority from which they derive.”<sup>270</sup> Darwall writes the following about the justification of “differential authority relations”, like those found in “legal and political philosophy”:

The lesson to draw... is that the only justification that can succeed is one that proceeds from within the second-person standpoint, beginning with the assumption that we all share a common basic authority to make claims and demands of one another at all, and proceeding from there to consider what differential claims to authority anyone could sensibly accept, or no one could reasonably reject on that basis.”<sup>271</sup>

While Darwall obviously has a great deal to say about authority in general, his work contains remarkably little discussion of the special case of political or legal authority. In the rest of the chapter, I will show that there are two different ways of extending Darwall’s general theory of second-personal authority to the specific case of political authority, each hinted at by remarks made in different parts of his work. I will argue that neither of them provides a satisfactory way for liberals to reconcile authority and autonomy.

## 5 Darwall’s First Theory of Political Authority

The first way to extend Darwall’s theory of authority to political authority is the most direct one. On this view, political authority is the same kind of individual authority at stake in ordinary

---

<sup>270</sup> Darwall, “Law and the Second-Person Standpoint,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 174. It is unclear whether Darwall intentionally shifted from explicit acceptance to hypothetical acceptance or reasonable acceptance. If it is intentional, perhaps he does so to account for those relationships of authority that are not so clearly accepted by both parties. (This is plausibly true in the case of Pablo and Abigail, if we imagine that Abigail refuses to acknowledge Pablo as an equal.) It is also possible that Darwall didn’t really change his mind, and instead intends for these to be two separate ways of justifying claims or demands made by individual authority. If that is indeed his view, individual authority would be justified (a) when two people freely accept a relationship of accountability that is *consistent* with their standing as members of the moral community (e.g. in the waiter and customer case), *or* (b) when two people should reasonably be expected to accept a relationship of accountability *given* their standing as members of the moral community (e.g. the case of Pablo and Abigail).

<sup>271</sup> Darwall, “Authority and Reasons: Exclusionary and Second-Personal,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 167.

rights and claims, like the kind found in the case of Pablo and Abigail.<sup>272</sup> Recall that Darwall defines authority (in general) in the following way:

Someone has practical authority with respect to another if, and only if, the latter has a second-personal reason to comply with the former's valid claims and demands and is accountable to the former for so doing.<sup>273</sup>

If we were to say that political authority is this kind of authority, it would involve the capacity to make demands on other people, a capacity which is rooted in a relationship of accountability and responsibility between the state and subject. As Darwall writes, "The concept of legal obligation seems analytically to entail those of legal responsibility, authority, and reasons. What one is legally obligated to do is what it is one's legal responsibility to do, what one has a legal reason to do, and what legal authority requires or demands that one do."<sup>274</sup> To understand the state as having authority in this sense, we need to make sense of how the state can have a *right* that someone do something, and how that person can be accountable to the state for so-doing. For example, then, when the state passes a law that says that the maximum speed that citizens can drive is 100km/hr, we should understand this as the state making a claim on citizens, demanding that they drive no more than 100km/hr, and that the state would be entitled to hold them accountable for doing so (through guilt, resentment, and so on).<sup>275</sup> Call this Darwall's *accountability condition*.

Citizens, in turn, would be under bipolar obligations to conform to the state's demands. Because the obligations involved in individual authority are bipolar obligations, if the state were to exercise individual authority over its subjects, the state's will must be a distinct source of

---

<sup>272</sup> I take it as straightforwardly true that political authority is clearly not representative authority; it does not seem to be exerted implicitly or necessarily in the way that representative authority is.

<sup>273</sup> Darwall, "Authority and Reasons: Exclusionary and Second-Personal," 266.

<sup>274</sup> Darwall, "Law and the Second-Person Standpoint," in *Essays in Second-Personal Ethics*, 173.

<sup>275</sup> The right involved in political authority might be understood as a cluster of many separate rights to make many different claims ("The state has a right to tell you to drive on the right hand side of the road", "the state has a right to tell you to drive on the left hand side of the road", etc...), or as a right to make one single open-ended claim ("The state has a right to be obeyed").

normativity for the person who is subject to them. It is not enough, in other words, for the state's will to be treated as a "feature of the moral landscape" that bears on the subject's moral obligations – the will of the state must be an "intrinsic source of bindingness in its own right."<sup>276</sup> Call this Darwall's *intrinsicity condition*.

We have already seen that the accountability condition and the intrinsicity condition would rule out the normal justification thesis as a substantive theory of political authority, but interestingly it appears that these conditions would rule out many other substantive theories of political authority as well. For example, fair play theories of authority, which hold that our obligation to obey the law is borne from a duty of fairness owed to other citizens, would also clearly be ruled out on this theory. This is because the obligation of fairness owed to other citizens does not regard the will of the authority as an intrinsically important feature of the situation. We can see this because the state need not be a moral agent for it to be fair or unfair that one obeys it as long as others are obeying it – it is irrelevant, in fact, whether it has a will at all.

## 6 A Solution to Wolff's Challenge?

Again, the reason that I have raised the second-personal account of authority in this thesis is that it offers the hope of being distinct from moralism. It seems like the second-personal is a broader category than the moral. Often when we address people, and make demands of them, we don't intend for anything moral to be going on, even though we expect that they are going to do what we have demanded them to do. Moreover, by distinguishing bipolar obligations and moral obligations, Darwall appears to have offered a way of explaining how second-personal obligations might have a distinct normative force from moral obligations. Finally, if these second-personal obligations are important enough, this offers the prospect of an explanation of why agents might be entitled to not act on moral considerations in pursuit of them.

---

<sup>276</sup> Darwall, "Bipolar Obligation," in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 29.

I want to emphasize that this is at most a sketch of a response to Wolff's challenge. There are many unanswered questions about it. Darwall's argument at best establishes that other people's choices, and the expressions of their wills, are sometimes sound objects of deliberation, things that can weigh on the reasons that we have to perform or not perform certain actions. To fully elaborate on this view, we would need to offer a more substantive explanation of *when* and *why* an expression of someone's will might make it the case that their listener does not need to act on the moral considerations that seem pressing to them. Most of all, we would need to explain why purely second-personal considerations might trump moral obligations, and when they can, why they do so in some cases but not others. For Darwall, accountability will play a role in this explanation, so we would also need some kind of explanation for when people are specifically accountable to one another. However, in the next section I will argue that any further elaboration is likely to be fruitless either way. Any attempt to treat political authority as second-personal will face the same problems that have faced the moralist theories discussed earlier in this thesis.

## 7 The Failure of the Solution

Let me begin with an initial worry about Darwall's view that is not especially decisive, but one that has occurred to some of the philosophers who have commented on it.<sup>277</sup> One might say that the accountability condition and intrinsicality condition involve an illegitimate personification of the state, even more than other theories of authority. It already sounds a little strange to talk about being accountable to the state; it certainly sounds strange to talk about the state *guilting* its citizens if they don't comply with its demands. However, I do not think this is the serious obstacle to a second-personal account of political authority. When we talk about the state having a will, or being capable of various sentiments, we can imagine the will of its leader, in their office of leader, or the cumulative will of the members of a parliament, or the cumulative will of the population as a whole. Instead, I think that the real problem with this first second-personal theory of political authority is that it collapses either into moralism or consent theory.

---

<sup>277</sup> For example, Enoch, "Authority and Reason-Giving," 325.

We can first glimpse this in Darwall's examples of second-personal authority. Many of Darwall's examples are moral ones, like his example of Pablo demanding that Abigail remove her foot off of his foot. The remainder of his examples involve consent, whether actual or hypothetical, like his example of the drill sergeant ordering the private, or the customer asking the waiter for a cup of coffee. We have already seen why an account of political authority that rests on moral theory or rests on consent will not suffice to address Wolff's problem, at least for a liberal.

Aside from this somewhat circumstantial evidence, there are two more theoretical aspects of his view that further show that there is a deep connection between political authority and morality. First, as we have already seen, Darwall seems to think that every "bipolar" obligation (a purely second-personal obligation, owed to one individual) still entails a moral obligation (an obligation that is owed to the whole moral community). His basic thought is that to be under a bipolar obligation, a person must be capable of blaming themselves, and to be capable of blaming themselves, that person would have to see themselves as blameworthy from the impersonal moral point of view, from the perspective of the moral community as a whole.

Second, Darwall thinks that all second-personal phenomena carry certain moral presuppositions about respect and agency; indeed, the ultimate conclusion of his book is that our second-personal involvement with one another presupposes some version of the Categorical Imperative which in turn is grounded in a kind of contractualism. Darwall writes, "When we acknowledge the summons of another free and rational agent... We presuppose the equal dignity of rational beings and our ability to act on a "law" or reason."<sup>278</sup> A version of contractualism "can be found in the commitments within the second-person standpoint."<sup>279</sup> Thus in order to accept the force of second-personal considerations, one needs to accept a certain form of Kantian morality to begin with. It would seem to follow that those second-personal considerations are not going to be able to trump Kantian or contractualist morality.

---

<sup>278</sup> Darwall, *The Second-Person Standpoint*, 33.

<sup>279</sup> Darwall, *The Second-Person Standpoint*, 37.

Since the force of second-personal considerations comes in part from their relationship to morality, it is hard to see how they would give commands any genuinely exclusionary force over an agent's moral reasons. An agent who was commanded to do something might find that they have a new second-personal reason to do what was demanded, but it is hard to see how that would make it the case that they no longer have to act upon their other, pre-existing moral reasons. It seems to be no different than the example of fair play mentioned earlier in this thesis, but where we substitute "the fact that someone demands it" with "fairness demands it".

## 8 Darwall's Second Theory of Political Authority

Darwall suggests a rather different theory of political authority when he writes that it is "truism" that a legal system at least claims to have "the authority to make or find law itself..."<sup>280</sup> From there, Darwall elaborates on how the second-person standpoint sheds light on the distinction between civil law and criminal law:

It is uniquely up to a(n alleged) victim of a violated bipolar obligation to decide whether or not to bring a case in the civil law of contracts or torts. If a wronged or injured party would prefer not to pursue a tort action seeking compensation, the state and other citizens do not generally have the authority to pursue it on her behalf. It is not, however, up to a(n alleged) victim to decide whether or not to pursue a criminal case, including for the very rights violation of which she has been victim. That is up to "the people" and their representatives. The criminal law is to the moral law as civil law is the dialogical [roughly second-personal] order to bipolar obligations.<sup>281</sup>

In this passage, Darwall draws a distinction between the obligations of civil law and the obligations of criminal law. He notes that parts of the civil law (like torts) involve violations of bipolar obligations owed to individuals, which are remedied through the victim demanding

---

<sup>280</sup> Darwall, "Law and the Second-Person Standpoint," in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 172.

<sup>281</sup> Darwall, "Bipolar Obligation" in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 31.

compensation that is owed to them. For example, if I cause an automobile accident in a negligent manner, the person whom I have harmed has a right to demand compensation from me, though they are not obliged to demand this compensation if they do not want to do so. The obligations of civil law, then, are analogous to bipolar duties. In contrast, Darwall notes that the criminal law is the part of the law for which punishment is warranted, and this punishment must be seen as appropriate from “a common perspective that violator and victim share as representative persons or members of the moral community”.<sup>282</sup> It is the whole community’s concern when a violent assault goes unpublished, and the decision of whether to punish the wrongdoer is not at the discretion of the victim. We can see, then, that the law contains analogues to both bipolar obligations and moral obligations.

Darwall does not himself explicitly say this, but when he describes legal authority as a power to “find” or “make” law, he must mean something different from the individual authority that is associated with bipolar obligations. This observation implies that political authority involves not just a right to make claims and demands, as in individual authority, but a power to create bipolar obligations that are owed to others, as well as moral obligations (or at least obligations analogous to moral obligations).<sup>283</sup> These obligations are owed to some person or body other than the state. In the case of civil law, the obligation is owed to the wronged party, like the victim in an automobile accident. In the case of the criminal law, the obligation is owed to the community as a whole. So according to this second view of political authority, political authority is the state’s power to create a variety of obligations among those subject to it, including obligations that are owed to fellow citizens rather than the state itself. It’s not just that

---

<sup>282</sup> Darwall, “Law and the Second-Person Standpoint,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 176.

<sup>283</sup> In many respects, Darwall’s second theory has many similarities to Stephen Perry’s moral power theory of authority in “Political Authority and Political Obligation”. Perry’s full definition of a moral power is the following:

One person A has a power to effect a certain kind of change in the normative situation of another person B if there is reason for regarding actions which A takes with the intention of effecting a change of the relevant kind as in fact effecting such a change where the justification for so regarding A’s actions is the sufficiency of the value or desirability of enabling A to make this kind of normative change by means of this kind of act. (Perry, “Political Authority and Obligation,” 25)

the state is making a claim that is appropriate for it to make, but that it is changing which claims are appropriate to make. In Hohfeldian terms, this second view equates political authority with a power rather than a right.

There are two crucial parts of this second view that Darwall leaves unclear. First, it is unclear whether this second view of political authority must also meet the accountability and intrinsicity conditions. As we saw earlier, Darwall thinks that bipolar obligations and moral obligations are pre-emptive, and that a power to create *any* kind of pre-emptive reason for action always requires a relationship of accountability in the background.<sup>284</sup> If so, this would lead to a somewhat strange view where laws entail obligations that are simultaneously owed to the state *and* to fellow citizens, or that subjects are accountable to the state for being accountable to others. Second, Darwall says essentially nothing about the justification of these relations – when it is that they come about. We can only rely on his earlier suggestion about differential authority relations: that “anyone could sensibly accept [them], or no one could reasonably reject [them]”, given our basic representative authority as moral agents.<sup>285</sup>

Nevertheless, as with the first theory, it is difficult to see how any elaboration of these details would help this theory to respond to Wolff’s challenge. If there was such a thing as a power to create new moral duties, it is hard to see how those newly-created moral duties would outright displace existing ones in the way that is characteristic of authority. It seems more likely that they would simply create new moral reasons of some weight, perhaps greater, perhaps lesser, than the ones that already exist.

## 9 Conclusion

Despite the initial hope that the second-person standpoint offers a genuine alternative to grounding political authority in moralism and practical reason, I have argued that it collapses

---

<sup>284</sup> Darwall, “Authority and Reasons: Exclusionary and Second-Personal,” 261.

<sup>285</sup> Darwall, “Authority and Reasons: Exclusionary and Second-Personal,” in *Morality, Authority and Law: Essays in Second-Personal Ethics I*, 167.

into moralism or consent theory. Once again, liberals will need to look elsewhere to find an explanation of why members of a society must not act what they take to be their moral duties, and why they should obey the state instead.

## Chapter 8 Conclusion

In this thesis, I have argued that liberal theory has failed to reconcile authority and autonomy. Neither morality, rationality, nor the intrinsic bindingness of other people's wills provide a standpoint from which it can make sense to authoritatively exclude moral considerations when determining how to act. This means that although many people, including many citizens of liberal states, believe that their states have legitimate authority over them, so far this belief appears to be without merit.

Since liberalism insists on the right to make up one's own mind about important moral matters, while legitimate authority would apparently seek to override that right, perhaps it is no surprise that liberalism and legitimate authority are incompatible with one another. However, I have tried to argue that this conclusion is far from obvious; to establish it, one must first consider the ingenious range of ways that liberalism has tried to preserve for the state some degree of exclusionary authority. Indeed, some of these views came closer to reconciling authority and autonomy than one might have expected. I also hope to have shown that the promise of each of these views is best appreciated in light of their relationship to simple moralism. Justice as fairness can be seen as a complex two-step kind of moralism. The service conception of authority is an attempt to go beyond moralism and ground authority in principles of rationality rather than directly on moral principles. And the second-personal theory of authority, at least in the way I have interpreted it here, tries to ground authority on the intrinsic normativity of interpersonal relationships.

While arguing that these theories are not successful, I also hope to have shown that the most important fault lines within the debate over political authority are not where many liberal political philosophers have located them. The most important divisions among liberals are not between consent theorists, fair play theorists and their peers, but instead concern the fundamental kind of reason that we have to obey the law. Is it a moral reason, a requirement of rationality, or

something else entirely?<sup>286</sup> This is the fundamental question that must be answered before liberals can give a specific, concrete theory of why the state enjoys legitimate authority over us.

Would it be so terrible for liberals to embrace Wolff's anarchism and accept that the state does not possess legitimate authority? Perhaps, in spite of the variety of moral views within a society, there is enough agreement between its members to coordinate on important matters without recourse to authority. For example, some have suggested that social coordination might be achieved without authority by simply highlighting the salience of a single choice for members of a society to converge upon. Wolff has even suggested that "the army itself could be run on the basis of voluntary commitments and submission to orders."<sup>287</sup> To the extent that certain social aims do need legitimate authority to be pursued, Wolff would say that those aims are not worth pursuing. He thinks that the belief in legitimate authority typically enables "the collective pursuit of some external national goal such as national defense, territorial expansion, or economic imperialism", or the "maintenance of our industrial economy".<sup>288</sup>

Moreover, just because liberals have to make do without Wolff's demanding sense of legitimate authority in terms of obeying the law 'just because it is the law', that doesn't mean that they can't endorse some other sense of legitimate authority. For example, the liberal might say that the state possesses legitimate authority in the sense of having a justification right to coerce its citizens. That is, perhaps when we say the state has legitimate authority, perhaps this means that the state has the moral entitlement to coerce its citizens into following the law, but that citizens have no corresponding moral obligation to obey the law. This would sidestep Wolff's concerns about autonomy, because these concerns are limited to an agent's voluntary choices.<sup>289</sup>

---

<sup>286</sup> It is also important to note that consent theory, fair play theory and associative obligation theory might be read in different ways. For example, I have frequently presented fair play theory as a simple moralist theory, but it need not be read in this way. Likewise, we do not necessarily need to regard the force of consent as coming from moral duties to keep promises – it might have some alternative ground.

<sup>287</sup> Wolff, *In Defense of Anarchism*, 80.

<sup>288</sup> Wolff, *In Defense of Anarchism*, 79.

<sup>289</sup> It may well turn out that legitimate authority in this alternative sense is not compatible with our autonomy either, but this is not established by Wolff's argument on its own.

Despite these attempts to soften the blow of doing without legitimate authority, I think that liberalism would suffer without it. Even if we accept Wolff's suggestion that we can do without complex social organization, anarchism undermines commonly-held intuitions about the state. As Leslie Green writes, part of the "self-image" of the state is that it claims to possess legitimate authority (and supreme legitimate authority, no less).<sup>290</sup> For this reason, those who are committed to legitimate state authority may still treat the argument of this thesis as a *modus tollens*: because liberals cannot account for the state's legitimate authority, liberalism must be mistaken.

That said, I do not claim to have produced a conclusive argument that it is impossible for liberal theory to reconcile authority and autonomy. There might be standpoints that I have not considered, or modifications of the views considered here, that might be more suitable for reconciling authority and autonomy. I will end this thesis with one brief suggestion offered in this spirit. This suggestion is drawn from Wolff himself, who suggests that it is permissible for an autonomous agent to submit to an *arbitrator's* authority as a way of resolving their disputes with other agents:

For example, a community may agree unanimously on some principles of compulsory arbitration by which economic conflicts are to be settled. An individual who has voted for these principles may then find himself personally disadvantaged by their application in a particular case. Thinking the principles fair, and knowing that he voted for them, he will (hopefully) acknowledge his moral obligation to accept their operation even though he would dearly like not to be subject to them .... this individual will have a moral obligation to obey the commands of the mediation board or arbitration council, whatever it decides, because the principles which guide it issue from his own will. Thus the board will have authority over him (i.e. a right to be obeyed) while he retains his moral autonomy.<sup>291</sup>

---

<sup>290</sup> Green, *The Authority of the State*, 86.

<sup>291</sup> Wolff, *In Defense of Anarchism*, 25.

At first glance, this suggestion seems radically out of step with his overall argument for anarchism. Given that argument, submission to arbitration should be impermissible for the same reasons that consent theory and authority more generally are impermissible. He does not elaborate on why he thinks that arbitration would fare any better than the other views of authority he dismisses in his book. That said, even if he does not supply the argument himself, perhaps he is onto something. It seems to me that arbitration occupies a strange middle ground between many of the theories considered here. It eludes easy characterization in terms of any of them taken individually. Disputants typically consent to arbitrators, but this seems like a necessary rather than sufficient condition for their authority. Nor is arbitration something that follows from the correct moral theory – indeed, it is what we consult when we *disagree* about morality.

The contemporary philosopher who has done the most to explicitly advance arbitration as a theory of authority is perhaps Daniel Viehoff. He says that the point of arbitration is to “settle disputes while upholding certain values central to the disputants relationship.”<sup>292</sup> These values relate to “non-subjection” – the importance of having certain interpersonal relationships whose parameters are not decided simply by the power advantage of their constituent parties.<sup>293</sup> To take an example from outside of political philosophy, we might imagine that two married people have reason to decide how they will live together (e.g. how their children will be educated, where they will go on vacation) without appealing to the power imbalances that arise between them because of differential earnings, gender, and so on.

While in certain cases, like private relationships, the parties can simply commit to not acting on the relevant power imbalances, Viehoff appears to suggest that this solution is often ineffective or unavailable in the case of citizens.<sup>294</sup> Citizens jostle with one another to decide how their society should be run, and the more powerful will typically get their way. This imbalance prevents them from enjoying a relationship of civic equality, something like being able to stare each other in the eye and know that they each have had an equal right to contribute

---

<sup>292</sup> Viehoff, “Debate: Procedure and Outcome in the Justification of Authority,” 255.

<sup>293</sup> Viehoff, “Debate: Procedure and Outcome in the Justification of Authority,” 255.

<sup>294</sup> Viehoff, “Democratic Equality and Political Authority,” 367

to the terms that govern their interactions. In this case, Viehoff suggests that citizens should allow the state to play the role of arbitrator in their disputes. When they face a dispute over how to coordinate their activities, by treating the law as an exclusionary reason, they exclude considerations involving power imbalances from their decision-making about what to do, and thus preserve the relationship of civic equality. This benefit is realized regardless of whether the law is correct on the merits, at least within limits. What seems to matter, instead, is that the law is appropriately fair – that the law did not intentionally favour one of the disputants over the other. Even if the law tells two citizens to follow a somewhat inferior coordination scheme, and they obey it, they still will have excluded the reasons that, when acted upon, undermine their relationship of civic equality.

We can see that Viehoff's view is closely related to the service conception of authority. Raz himself sees arbitration as a clear-cut example of the service conception: an arbitrator's authority hinges on her ability to better appreciate the reasons that apply to disputants, through having greater knowledge, being more impartial, or being better positioned to initiate coordination.<sup>295</sup> In effect, Viehoff accepts the thought that authorities help subjects act on their reasons, but he enlarges the relevant set of reasons. On Viehoff's view, those reasons include not only the reasons that speak in favour or against an action, but also reasons that have nothing to do with the commanded action itself, like the reasons they have to uphold equality within certain relationships.<sup>296</sup>

Viehoff's view offers some initial hope for addressing Darwall's objection to the service conception. According to him, an arbitrator's directives are not exactly like advice, because their bindingness does not strictly speaking depend on whether they are correct or incorrect, so he

---

<sup>295</sup> Raz, *The Morality of Freedom*, 41.

<sup>296</sup> To some degree, as we saw in chapter 5, I think Raz already accommodates this enlargement of the set of reasons. This seems to be present in his examples of cases where authorities help subjects choose between two indistinguishable alternatives, and his examples of cases where authorities can act on the reasons that apply to them alone (e.g. *de minimis* constraints). So there Viehoff's view and Raz's view may be more similar than Viehoff intends for them to be.

cannot be accused of mixing up ‘counsel’ and ‘command’ in the way that Raz appears to do.<sup>297</sup> Nevertheless, I think there is good reason to worry that Viehoff’s view still falls prey to Darwall’s objection. Even if two people have a valuable relationship, and there is an arbitrator who can help them to manage their relationship without recourse to their power dynamics, this doesn’t entail that the arbitrator actually has authority over them. It is akin to the human alarm clock example discussed earlier in the thesis, but instead of helping a subject act on their own reasons, the arbitrator helps a pair of subjects act on their reasons to have certain kinds of relationships. There is one difference between arbitration and the human alarm clock example that might be important: Viehoff’s account of authority is over pairs of individuals or groups rather than individuals. However, without further argument, it is unclear how this would make a difference from the point of view of Darwall’s objection.<sup>298</sup> Thus I suspect that a successful account of political authority as arbitral authority must distance itself even further from Raz’s service conception of authority if it is to find a way of resolving Darwall’s objection (and Wolff’s challenge more generally).

---

<sup>297</sup> Note that to the extent this is also true of Raz’s theory, as suggested in the previous footnote, then the service conception of authority likewise escapes this part of Darwall’s criticism.

<sup>298</sup> Since arbitration occurs over multiple individuals, this raises the possibility that each individual might have a right that their dispute be arbitrated. This would introduce the concept of right back into authority, which seems like a good start for addressing Darwall’s objection, but it would be the wrong party who has the right: the disputants would have rights against one another, but the arbitrator would not have a right against the disputants.

## Bibliography

- Alexander, Larry. "Law and Exclusionary Reasons." *Philosophical Topics* 18, no. 1 (1990): 5-22.
- Beauchamp, Tom and Ken Witowski. "A Critique of Pure Anarchism." *Canadian Journal of Philosophy* 2, no. 4 (1973): 533-539.
- Darwall, Stephen. "Authority and Reasons: Second-Personal and Exclusionary." *Ethics* 120, no. 2 (2010): 257-278.
- Darwall, Stephen. *Morality, Authority and Law: Essays in Second-Personal Ethics I*. Oxford: Oxford University Press, 2013.
- Darwall, Stephen. *The Second-Person Standpoint*. Harvard: Harvard University Press, 2009.
- Dworkin, Gerald. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press, 1988.
- Dworkin, Ronald. "Lord Devlin and the Enforcement of Morals." *Yale Law Journal* 75 (1967): 986-1005.
- Dworkin, Ronald. "The Original Position." *University of Chicago Law Review* 40, no. 3 (1973): 500-533.
- Enoch, David. "Authority and Reason-Giving." *Philosophy and Phenomenological Research* 89, no. 2 (2014): 296-332.
- Frankfurt, Harry. "The Anarchism of Robert Paul Wolff." *Political Theory* 1, no. 4 (1973): 335-341.
- Green, Leslie. *The Authority of the State*. Oxford: Clarendon Press, 1988.
- Hart, H.L.A. *Essays on Bentham*. Oxford: Oxford University Press, 1982.
- Hershovitz, Scott. "The Role of Authority." *Philosophers' Imprint* 11, no. 7 (2011): 1-19.
- Hobbes, Thomas. *Leviathan*. Indianapolis: Hackett Publishing Company, 1994.

- Ladenson, Robert. "Legitimate Authority." *American Philosophical Quarterly* 9, no. 4 (1972): 335-41.
- Langvaten, Silje. "Legitimate, but unjust; just, but illegitimate: Rawls on political legitimacy." *Philosophy and Social Criticism* 42, no. 2 (2016): 132-153.
- Perry, Stephen. "Second-Order Reasons, Uncertainty, and Legal Theory." *Southern California Law Review* 62 (1989): 913-994.
- Perry, Stephen. "Political Authority and Political Obligation." In *Oxford Studies in Philosophy of Law (volume 2)*, edited by Leslie Green and Brian Leiter, 1-74. Oxford: Oxford University Press, 2013.
- Pritchard, Michael S. "Wolff's Anarchism." *Journal of Value Inquiry*, 7, no. 4 (1973): 296-302.
- Rawls, John. *A Theory of Justice*. Cambridge: Harvard University Press, 1999.
- Rawls, John. *Collected Papers*. Harvard: Harvard University Press, 1999.
- Rawls, John. *Political Liberalism*. New York: Columbia University Press, 2005.
- Raz, Joseph. *The Authority of Law (Second Edition)*. Oxford: Oxford University Press, 2009.
- Raz, Joseph. "Authority, Law and Morality." *The Monist* 68 (1985): 295–324.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Books, 1986.
- Raz, Joseph. "Respect, Authority and Neutrality – a Response." *Ethics* 120, no. 2 (2010): 279-301.
- Raz, Joseph. *Practical Reason and Norms*. Oxford: Oxford University Press, 1999.
- Raz, Joseph. *Between Authority and Interpretation*. Oxford: Oxford University Press, 2009.
- Raz, Joseph. "The Possibility of Partiality." Unpublished draft.
- Raz, Joseph. "Facing Up: A Reply." *Southern California Law Review* 62 (1989): 1153-1235
- Regan, Donald. "Authority and Value: Reflections on Raz's Morality of Freedom." *Southern California Law Review* 62 (1989): 995-1096.
- Reiman, James. *In Defense of Political Philosophy*. New York: Harper and Row, 1972.

- Riley, Patrick. "On the "Kantian" Foundations of Robert Paul Wolff's Anarchism." *Nomos* 19 (1978): 294-319.
- Ross, W.D. *The Right and the Good*. Indianapolis: Hackett Publishing, 1988.
- Rossi, Enzo and Matt Sleat. "Realism in Normative Political Theory." *Philosophy Compass* 9/10 (2014): 689-701.
- Sciaraffa, Stefan. "On Content-Independent Reasons: Its Not in the Name." *Law and Philosophy* 28 (2009): 233-260.
- Shapiro, Scott. "Authority." In *The Oxford Handbook of Jurisprudence and Philosophy of Law*, edited by Jules L. Coleman, Kenneth Einar Himma, and Scott J. Shapiro, 384-439. Oxford: Oxford University Press, 1999.
- Simmons, A.J. *Moral Principles and Political Obligations*. Princeton: Princeton University Press, 1979.
- Smith, M.J.E. "Is there a Prima Facie Duty to Obey the Law?" *The Yale Law Journal* 82, no. 5 (1973): 950-76.
- Sterba, James P. "The Decline of Wolff's Anarchism." *Journal of Value Inquiry* 11, no. 3 (1977): 213-217.
- Taylor, Charles. *Philosophical Papers*. Cambridge: Cambridge University Press, 1985.
- Thomas, Judith Jarvis. "A Defense of Abortion." *Philosophy and Public Affairs* 1, no. 1 (1971): 47-66.
- Viehoff, Daniel. "Debate: Procedure and Outcome in the Justification of Authority." *The Journal of Political Philosophy* 19, no. 2 (2011): 248-259.
- Viehoff, Daniel. "Democratic Equality and Political Authority." *Philosophy and Public Affairs* 42, no. 4 (2014): 337-375.
- Weber, Max. *The Theory of Social and Economic Organization*. Eastford: Martino Fine Books, 2012.

Wenar, Leif. "John Rawls." In *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), edited by Edward N. Zalta, <https://plato.stanford.edu/archives/sum2021/entries/rawls/>

Williams, Bernard. *In The Beginning Was the Deed*. Princeton: Princeton University Press, 2005.

Wolff, Robert Paul. *In Defense of Anarchism*. New York: Harper and Row, 1970.

Wolff, Robert Paul. "Reply to Prichard and Smith." *The Journal of Value Inquiry* 7, no. 4 (1973): 303-306.

Wolff, Robert Paul. "Reply to Reiman." <http://www.ditext.com/wolff/reiman.html>

