Ulysses Rebound: Recommitting Constitutional Precommitment

Christina Rothwell

A Dissertation submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Philosophy, York University, Toronto, Ontario

May 2025

Abstract

Standard arguments around constitutional precommitment – both in favour and against – presuppose a mistaken view about the concept of commitment itself. I will call this mistaken belief the *agent-dependent condition*. The agent-dependent condition is a feature of the operative understanding of commitment which requires that commitment, to be real or true or valid, be imposed on oneself *by oneself*. *Real* commitments, on this view, are necessarily *self*-imposed. To me, constitutionalism's significant focus on democratic legitimacy arises because *democracy* serves as a conceptual shorthand for the *agency* which is thought necessary in the concept of commitment: a commitment which has not been chosen seems no commitment at all. This is the implication of the agent-dependent condition.

Conflating the concept of commitment with the agent-dependent condition is a mistake, the pervasiveness of which is a result of a gross misrepresentation of the real core of commitment. I want to challenge the agent-dependent condition of commitment, arguing that whether and how one is committed is often a matter quite outside of an individual's control. I put forward a novel conception of commitment that places less emphasis on the agent and more emphasis on the conceptual constraint of commitment. In doing so, I choose to present these ideas from within and alongside the constitutional precommitment debate for two reasons. First, the two most general objections to constitutional precommitment – the 'non-bindingness problem' and the 'binding-others problem' – rely on the very confusions that make up the core of the agent-dependent conception of commitment. As such, the constitutional precommitment debate provides a useful frame in which to situate and demonstrate my arguments.

Second, my conceptual work on commitment may serve the constitutional debate by allowing for an alternate means of discussing the purpose and rules of the constitutional game, thereby offering some theoretical wiggle room. The tools I develop in my conceptual work on commitment can be used to develop a view of constitutional precommitment motivated by the concept of commitment itself. This is something far enough 'outside the box' to provide some novel questions and ideas for those working on constitutional precommitment, and constitutionalism more broadly.

Table of Contents

Chapter 1: Introduction

In 1979, philosopher Jon Elster made what he considers to be a mistake.

In his work *Ulysses and the Sirens: Studies in Rationality and Irrationality* Elster suggests that state constitutions – and some of their features, like periodic elections – could be considered "the electorate's method of binding itself and of protecting itself against its own impulsiveness."[1] Elster's purpose in *Ulysses and the Sirens* was to argue that the human need for techniques of 'indirect rationality' – of which precommitment is one – is not incompatible with the belief that we are rational or have rational agency. His intention was to plant the seeds of a theory of rationality that recognises the reality of 'weak will' and humanity's ways of working around and against it. Elster only introduces the topic of constitutional precommitment – an already well-established view – as a tentative example of this imperfect rationality on a general scale.

This idea of constitutions as tools of rational precommitment – commonly called the "precommitment view" – is hardly unique to Elster. American founding father Alexander Hamilton, in letter 78 of the *Federalist Papers*, repeatedly aligns the "will of the people" with the Constitution. John Potter Stockton, former American Senator, declared constitutions to be "the chains with which men bind themselves in their sane moments that they may not die by a suicidal hand in their frenzy."[2] Samuel Freeman describes the constitution as a "rational and shared precommitment among free and equal sovereign citizens" that prevents them "from later changing their minds and deviating from their agreement and commitment to a just

[1] Jon Elster, *Ulysses and the Sirens*: *Studies in Rationality and Irrationality* (Cambridge University Press. 1979), 90.
[2] *Congressional Globe, 42nd Congress, 1st Session 574 (1871). Accessed at:* https://memory.loc.gov/cgi-bin/ampage

constitution."[3] Friedrich Hayek describes a constitution as an instance of Peter Sober binding Peter Drunk, writing that a constitution "does not involve an absolute limitation of the will of the people but merely a subordination of the immediate objectives to long-term ones."[4] The precommitment view, in short, is one that holds a significant following.

And yet, Elster eventually abandons his earlier statement about constitutions being tools of precommitment, partly due to the concerning objections against such an idea.

> In *Ulysses and the Sirens* I noted that the Ulysses metaphor for constitution-making is only partially valid, and in particular that the idea of society binding "itself" is a controversial one. Yet I do not think I fully understood the extent of the disanalogy between individual and collective self-binding. As in many other cases, the transfer of concepts used to study individuals to the behaviour of collectivities, as if these were individuals writ large, can be very misleading. For one thing, constitutions may bind others rather than being acts of self-binding. For another, constitutions may not have the power to bind in the first place.[5]

While the concern about the 'non-bindingness' of constitutions is partly created by Elster's particular conception of precommitment, the concern that constitutions bind others is already a significant topic in constitutional theory. This concern, which I'm calling the "binding-others" problem, is particularly salient for constitutional democracies, where the combined principles of constitutionalism and democratic self-governance have been argued to be in tension: 'we the people' cannot be self-governing if we are chained by a constitution not of our making.[6] Many different iterations of such arguments have been given in rejection of the precommitment view – such as the particular claim that it is undemocratic – but these particular

---

[3] Samuel Freeman, "Constitutional Democracy and the Legitimacy of Judicial Review," *Law and Philosophy* 9, no. 4 (1990): 354, http://www.jstor.org/stable/3504771.
[4] Friedrich Hayek*, The Collected Works of F. A. Hayek.* Vol. XVII, *The Constitution of Liberty: The Definitive Edition* ed. Ronald Hamowy (Routledge. 2011), 268.
[5] Jon Elster, *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints* (Cambridge University Press, 2000), 92.
[6] What this means will be explored in greater detail in Chapter 2.

arguments are all in service of the two general issues just noted: constitutions, if they can even bind at all, serve only to bind others.

Thus, by the time of *Ulysses Unbound*, Elster abandons his earlier statement, calling it a mistake.

> It is easy… to indulge in unsupported functional explanation. In *Ulysses and the Sirens*, which was a crusade against this mode of explanation, I made exactly this mistake when I argued that the system of periodical elections "can be interpreted… as the electorate's method of binding itself and of protecting itself against its own impulsiveness." Obviously, no electorate ever did anything of the kind.[7]

In my view, Elster *did* make a mistake, though not the mistake he has in mind. While constitutional precommitment is and has been controversial, I think Elster was on to something in his attempt to consider whether and how constitutional precommitment might work given his thoughts on (pre)commitment.[8] Typical arguments around constitutional precommitment (and whether it's theoretically possible or desirable) tend to focus on the question from the perspective of democratic theory or constitutionalism itself. Elster is, as far as I know, the only person to consider constitutional precommitment from within the concept of commitment instead. As such, I think Elster's initial steps were a great way forward. His *real* mistake was that he became too blinded by his theory of individual self-binding, confusing the conditions of individual rationality for the limits of the concept of (pre)commitment. When confronted with the fact that his model of individual self-binding may be incompatible with constitutional 'self'-binding, Elster seems to have concluded that he must choose either to revise his views of the former or abandon the latter; this was Elster's mistake.

As I will argue throughout the remainder of this work, it is my view that the standard arguments around constitutional precommitment – both those in favour and those against –

---

[7] Elster, *Ulysses Unbound*, 90.

[8] Elster uses commitment, precommitment, and self-binding interchangeably. I use this same convention in these discussions, though I do point out the problem of conflating commitment and self-binding.

presuppose a mistaken view about the concept of commitment itself. I will call this mistaken belief the *agent-dependent condition*. The agent-dependent condition is a feature of the operative understanding of commitment which requires that commitment, to be real or true or valid, be imposed on oneself by oneself; constraint from any other source is not true or real commitment but is constraint of some other kind. *Real* commitments, on this view, are necessarily *self*-imposed.

The agent-dependent condition is clearly present in Elster's work, insofar as the relation between precommitment and individual agency was precisely his focus. But the condition is also clearly adopted in the constitutional precommitment debate through the extent to which arguments from *both* advocates *and* critics of constitutional precommitment centralise the importance of democratic legitimacy.

When critics of constitutional precommitment reject the view, their most significant problem is (what is called) 'the democratic worry': the view that purported constitutional commitments are often *not* chosen by those over whom they are meant to be binding and are thus democratically illegitimate. When advocates of constitutional precommitment respond to the democratic worry, they attempt to do so by demonstrating how those constitutional commitments *have* been chosen – albeit in roundabout ways – and therefore the democratic worry is no more. To me, the significant focus of democratic legitimacy here arises because *democracy* serves as the conceptual stand-in for that *agency* which is thought necessary in the concept of commitment: a commitment which has not been chosen seems no commitment at all. This is the implication of the agent-dependent condition.

Of course, we could also say that the relevance of agency is easily sourced from the concept of democracy, since to say something is democratically legitimate is perhaps just to say

that the relevant self – in this case, the *demos* – has chosen it for itself. Whatever the conception of commitment, it would seem like democratic legitimacy still requires that the people choose commitments for themselves; "by the people, for the people" after all. But though I can appreciate the ease with which we may connect the concepts of agency and democracy, I think doing so is a distraction.

Despite the amount of work done on precisely the issue of whether constitutions can (or do) adequately represent the *demos*, there is a relative lack of work done on the concept of commitment itself. Elster's work is arguably the most influential conceptual work on precommitment, partly because it seems to be one of the only works in the area. But for better or for worse, Elster's work simply isn't up to the task of dealing with a theory like constitutional precommitment, specifically because his model was exclusively and deliberately limited to dealing with individual agency; he simply doesn't consider commitment outside of the very narrow perspective of the individual rational agent. As such it seems that, after Elster, the concept of commitment itself has come to mean nothing more than an agent-initiated constraint.

To conflate the concept of commitment with the agent-dependent condition is a mistake, the pervasiveness of which is a result of a gross misrepresentation of the real core of commitment and, I think, a general discomfort with constraint and lack of agency. I want to reconsider these discomforts by challenging the agent-dependent condition of commitment, arguing that whether one is committed is a matter quite outside of an individual's control. Most significantly, I do not mean to fully sever the perceived connection between our values and our commitments. While undoubtedly the things we value and choose for ourselves matter for making sense of how we deal with the commitments we do have, centring on the agent and their values results in an incomplete view of commitment.

I put forward a novel conception of commitment that places less emphasis on the agent and more emphasis on the conceptual constraint of commitment. In doing so, I choose to present these ideas from within and alongside the constitutional precommitment debate for two reasons. First, it seems that the two most general objections to constitutional precommitment – the 'non-bindingness problem' and 'binding-others problem' – rely on the very confusions that I think make up the core of the agent-dependent conception of commitment. As such, the constitutional precommitment debate provides a useful frame in which to situate and demonstrate my arguments, which would otherwise consist of several highly theoretical and abstract statements and observations. The constitutional precommitment debate is thus a very useful illustration of how things go wrong and how things might look otherwise.

Second, I think my conceptual work on commitment may serve the constitutional debate by allowing for an alternate means of discussing the purpose and rules of the constitutional game, thereby offering some theoretical wiggle room. It seems to me that, post-Elster, the precommitment view has been relatively stagnant. Since the issue of constitutional precommitment has paused at the question "is constitutional precommitment democratically legitimate," most of the novel arguments in this area focus on theories of democratic self-governance, and ultimately end up doing little more than advancing one or another conception of democracy. Elster's "mistake" was thus quite unique, precisely because it came at the constitutional precommitment issue from the perspective of commitment rather than from the perspective of democratic theory. Elster erred only in focusing too much on *individual* precommitment, but his original idea is, I think, worth pursuing. The tools I develop in my conceptual work on commitment can be used to develop a view of constitutional precommitment motivated by the concept of commitment itself. This angle may be far enough 'outside the box'

to provide some novel questions and ideas for those working on constitutional precommitment (and constitutionalism more broadly).

Methodology and Chapter Breakdown

Given the dual task of conceptual analysis and theoretical application, this dissertation will follow two general threads. One thread deals with the constitutional precommitment debate and follows the particular arguments about democratic legitimacy and the issues of constitutionalism to eventually get us to the issue of commitment. The other thread, then, deals with the concept of commitment itself (and, by extension, precommitment).

Chapter 2 introduces the constitutional precommitment debate by providing a simple discussion of both the intended purposes of constitutions and some history about how the precommitment view (and the existing objections to it) has come to be. Questions of democratic legitimacy are a common focus for those working on constitutionalism, as the tension between constitutions and democratic self-governance is thought to be relatively unique. Under the sort of 'standard' interpretations, democracy, legitimacy, and (it is believed) even commitment, all centre around the ability to choose. Constitutionalism seems to directly challenge these concepts, as the very purpose of a constitution is to constrain political power by limiting the set of possible political and legal choices. The precommitment view is an attempt to allay concerns about constitutional coercion by recasting imposed constitutional constraint as collective rational self-binding; under this framework, constitutional constraints shift from 'coercive' to 'commitment.' The constitutional precommitment view is thus meant to be a sort of salve to these democratic-minded concerns (aptly called 'the democratic worry'), but instead seems only to have furthered these concerns, giving rise to the argument that if constitutions manage to bind anyone at all, they manage only to bind others. Chapter 2, then, introduces us to the first general problem in constitutional precommitment: the 'binding-others' problem.

Chapter 3 introduces us to Elster's specific view of precommitment, which is necessary for making sense of the second general issue of constitutional precommitment: the 'non-bindingness' problem. Elster's intention in *Ulysses and the Sirens* is to outline the conditions under which self-binding 'counts' as a kind of indirect exercise of rational agency. As such, there are certain aspects of his view that are, I think, accidentally or illegitimately imported when he considers what 'constitutional precommitment' might look like. The 'non-bindingness' objection to constitutional precommitment rests on a very particular understanding of what kind of constraint a commitment must be to be a 'proper' Elsterian practice of agency; it will thus be useful to have some sense of what Elster thinks these exercises of agency must look like.

The main takeaway from this first thread is that there is a strongly held belief that constitutional constraints must be self-imposed (i.e. chosen by the demos) to be 'real' or 'valid' commitments. While it is easy to attribute this assumption to the conceptual tenets of democratic theory, the discussions of constitutional precommitment from the purview of democratic theory provide nothing more than frameworks of belief for reflective equilibrium. Agent-dependence may be a condition of democratic theory, but even this is a statement subject to interpretation. Regardless, the move from what *might* be the case for democratic theory to what *is or must* therefore be the case for the concept of commitment - and for constitutional precommitment - itself is mistaken. Bogged down by the agent-dependent condition, these standard arguments about constitutional precommitment seem unable to get past issues of whether the people have *chosen* to be committed and fail to consider what is the actual connection between (pre)commitment and choice. Chapter 3, then, will end with a pause on constitutional theory as we turn to discuss the concept of commitment.

Chapter 4 introduces the second thread of the dissertation with a discussion of existing work on commitment (of which there is little). The chapter begins with some observations about (what we might call) the 'folk' conception of commitment – 'commitment' as we speak of it in ordinary, day-to-day language – before turning to the views of Sam Shpall. Shpall's work on distinguishing moral and rational commitment is, thus far, the only focused conceptual work on commitment that I have found; all other discussions of commitment in academia presuppose some 'folk' understanding and treat the concept only instrumentally. Sam Shpall, much to my delight, is interested in the concept of commitment itself. What his discussion lacks in mention of constitutional law, it makes up for in its exclusive focus on the concept of commitment. While there is much in Shpall that I find valuable, I do take issue with one of his fundamental assumptions about commitment: it is taken for granted that commitments are normative constraints that an agent places on oneself – the self is the source of commitment. In other words, Shpall's theory of commitment presupposes the agent-dependent condition. I will argue that Shpall's reasons for thinking that commitment must be agent-dependent rest on mistaken beliefs about how commitment relates to other concepts (like requirement), and that once we properly review the concept of commitment, there is no longer any need for or benefit to having the agent-dependent condition.

In Chapters 5, 6, and 7 I turn to my own framework of commitment which I explain through a series of different distinctions and definitions. In Chapter 5, fresh from the discussion of Sam Shpall's work, I outline what *I* think are the sources of commitment, collectively called "the two-fold distinction." Unlike Shpall (and, to an extent, Elster) who thinks that individual agents are the sources of commitment, I think there are two sources of commitment: natural commitment, which is commitment by the way of the world around us; and social commitment,

which is commitment by human conceptual activity. The important difference – which I repeatedly highlight – between my view on the sources of commitment and the standard agent-dependent view is that on my view, commitment exists independent of any individual agent or their activity. Individual agency has less to do with whether one is committed, and more to do with how one deals with the commitments one nevertheless has.

Chapter 6 deals with several important concepts related to how exercises of agency fit into a model of commitment that appears indifferent to human agency. The first important concept is a posited concept of legitimacy. On this model, the word 'legitimacy' is reserved for the evaluative term used to categorise possible choices relative to the ends about which we reason: it does not presuppose any inherent acceptability, moral or otherwise. To be very clear about how I use 'legitimacy,' I spend some time outlining the characteristics I take the concept to have and differentiate my use of the word from how it is used in other areas (like political philosophy, for instance).

In addition to 'legitimacy,' Chapter 6 outlines what I call "the three-fold distinction", which refers to the three ways our choices relate to our commitments: completion (or satisfaction), breaking, and undoing (or dissolving). Whether we are committed is not up to us, as we are committed because of the way of the natural or social world around us. But we *do* have three different ways of engaging with the commitments we have, and our choices often operate at the level of choosing when and whether to complete, break, or undo certain commitments. Thus, the three-fold distinction (along with the concept of legitimacy) may be thought of as outlining how agency relates to commitment: given the particular circumstances of an individual's network of commitments (or, commitment matrix) there are differences in what different people

may legitimately do or not do, contingent on how different natural and social commitments bear on our options, decisions, and decision-making processes.

Chapter 7 deals with what I call 'forms' or 'types of commitment'. A form of commitment is a formal or structural commitment pattern that we can recognise, follow, and reinforce in many of our goals or collaborations. They are, in a sense, commitment games; the forms have certain rules and conditions that apply to all commitments of that type. I focus only on three types of commitment – resolution, fidelity, and deference – and endeavour to highlight how these structural rules might complicate the appearance of agreement on our decision-making processes when we are not clear about the nature of our commitments. The 'goals' of different agents may be described in the same words but rely on different forms of commitment and thus result in different assessments about what is or is not legitimate for each agent to pursue. This 'appearance' of agreement is an important point to notice, as it seems to be the culprit in a number of particularly tense philosophical discussions.

In Chapter 8, the two threads of the dissertation are united, and the discussion returns to the two problems of constitutional precommitment from the beginning: constitutions either don't bind at all (non-bindingness problem), or else they serve only to bind others (binding-others problem). With the tools I have outlined through the dissertation, I argue for an alternate perspective on the function of constitutions based on a revised understanding of the kind of constraint that commitment provides, and reconsider what the relationship between constitutionalism and democratic theory must look like, given the commitments of both.

My response to the non-bindingness problem rests on my concept of legitimacy and the three-fold distinction. The mistaken belief underpinning the non-bindingness problem comes (in part) from a very narrow understanding of the constraint of commitment. The constraining power

of commitment (and by extension, constitutions) is interpreted far too literally, and the fact that a constitution cannot actively prevent any individual from acting as they wish to act is taken to be a sign of its insufficiency as a constraining device. But the kinds of constraints that constitutions introduce are conceptual, not physical: constitutions are significant, and to overlook the contents of a constitution is, by extension, significant. While constitutions certainly don't bind us fast like Ulysses to the mast, the existence of a constitution does nevertheless change the normative landscape against which relevant actors act. The constraining power of a constitution is not that the paper itself may bind our hands, but that one who acts contrary to the constitution is seen as having done something significant and serious – not just to those who share the same constitution, but also to others who have constitutions of their own. While unconstitutional action might hold a different significance for fellow citizens than it might for allied countries or trade partners, there are relatively few instances where unconstitutionality is seen to be insignificant or ignorable by other players in the constitution game.

My response to the binding-others problem rests on the fact that it is not the objection it takes itself to be. On the agent-dependent condition model, it would certainly be a damning problem if it were the case that purported commitments were never actually chosen by those over whom they're meant to be binding: by definition, they could not be real commitments at all. Whatever the kind of conceptual work being done by others in this area, it seems equally taken for granted that the only 'real' commitments are those which are chosen for oneself. The agent-dependent condition thus transcends any particular theoretical project: it's not unique to any one conception of democracy, for instance. But the agent-dependent condition seems to me to do little more than try to alleviate concerns about coercion and constraint that operate in the background. Choice and agency are often relied upon as a sort of moral magic, but whether and

how they change the normative landscape is unclear at best (and absent at worst). For Elster, the purpose of discussing agent-dependent precommitment was clear, given his aim of discussing the conditions under which the passivity of precommitment is compatible with the activity of agency. But for those doing other conceptual projects, the distinction between commitments and requirements, or commitments and constraints, seems to me remarkably arbitrary.

On my framework, however, this is a non-issue; whether we are committed has nothing to do with whether we have chosen to be. Democratic legitimacy is less argumentatively significant for constitutional precommitment than is often thought, since the agent-dependent condition for commitment drops out. There's nothing about being committed that *requires* choice, and thus whether the *demos* have chosen the constraints that bear on it is irrelevant for the question of whether it is or can be committed. While agent-dependence may yet matter for arguments about *good* or *desirable* government, the democratic status of a constitution is irrelevant for matters of precommitment itself.

While such a conclusion could be written off as simply anti-democratic, I think there is value in recognising where energy is could be better expended. I personally consider and intend my conclusions to be reorienting rather than dismissive. The line of argument about whether the people have chosen or not is moot and played out; instead, we may do better to start from the observation that – no matter what – the people have commitments in which they have no choice, and yet are nevertheless required to make one. This line of thinking runs somewhat alongside more 'natural law' worldviews, but such abstract and global normative frameworks aren't necessary.

What I think will be clearly demonstrated by my dissertation is the fact that current debates about constitutional precommitment and its purported democratic legitimacy have, in

fact, nothing to do with constitutions or commitment. Instead, the topic is little more than a battleground for different camps to defend their respective views of what democracy is and requires. But constitutions are *not* democratic; that is precisely their purpose. If we're going to take constitutional precommitment seriously, it requires a real, clear-eyed recognition that we do not choose whether we are committed: good constitutional practice can only begin once we recognise this fact.

Chapter 2: Constitutionalism, Democracy, and Precommitment

Constitutional precommitment as a framework – compared to constitutionalism's full history – is relatively new. While recognisably constitutional practices have existed since ancient times, the theory of constitutions as precommitment devices – 'the precommitment view' – has developed only recently and as mentioned in Chapter 1, is both popular and controversial. The precommitment view is intended to allay concerns about the apparently unique tension between the principles of constitutionalism and the principles of democratic government,[9] but seems to create more problems than it solves. The two main issues for the precommitment view are, first, that constitutions may not actually commit anyone to anything, and second, that if they *do* commit, they can seemingly only commit others.

This chapter puts forward a brief outline of these issues, with particular focus on the issue of constitutions binding others.[10] My purpose here is not to argue for one conclusion or the other vis-à-vis whether constitutional precommitment is democratically legitimate, though I do have my view. Instead, I want to highlight the role that agency is given by both critics *and* advocates of constitutional precommitment alike.

Despite widespread disagreement on just about every other aspect of the discussion, advocates *and* critics agree that constitutional precommitment *must* be democratically chosen; they simply disagree on whether this condition is met. While one obvious source of this condition is the association between Democracy (of any kind) and agency, I wager that there is an additional source. I think there is an assumption that commitments must be chosen by oneself

[9] I say 'apparently unique,' because I am not convinced that democracy is the only social arrangement that exists in conflict with constitutionalism. Monarchies, it would seem, would also take issue with constitutional arrangement, since it would result in a purported sovereign leader who nevertheless must obey some other set of law. As an example of such an argument, see John Austin's *The Province of Jurisprudence Determined.*

[10] The issue of non-bindingness will be given more discussion in the next chapter.

to be legitimately binding on oneself; when unchosen, so called 'commitments' are nothing more than constraint or coercion.

Principles of Constitutionalism[11]

> Like a great work of architecture, a constitution combines form and function to produce lasting normative and aesthetic effects that transcend its component parts. Designed by the few and built by the many, it provides a space for expressing a society's values and expectations, and for imagining its future. It is both a powerful symbol and a symbol of power.[12]
> -B. MacLachlan

> As our country's supreme law, the Canadian Constitution sets out the basic principles of our democracy. It reaffirms our federal system of government, establishes courts for the administration of justice, protects our human rights through the Canadian Charter of Rights and Freedoms, and sets out the Aboriginal and treaty rights of Canada's first peoples.
> In sum, the Constitution is a living expression of who we are and what we value.[13]
> -D. Johnson

> The central notion of constitutionalism is that, although government is established to serve popular interests, government must be constrained in its powers if it is to retain authority in exercising those powers.[14]
> -A. Hutchinson

Put simply, constitutionalism is the study of constitutions: what they are, why we have them, and what it means for a state to be a constitutional state. Generally construed, a constitution is a legal instrument that outlines the fundamental structure of society. They are the highest law of the land, containing the rules determining whether and how to divide the powers of government, as well as the extent to which the state's power is limited.[15] Many constitutions contain charters or bills of rights, which usually outline some of the protected areas by which state powers are subject to limitation. To safeguard against their abuse, many countries'

---

[11] Portions of the wording in this Chapter are taken from or strongly reminiscent of C. Rothwell, *The Constitution of Theseus: Metaphysics of Constitutional Precommitment,* MA thesis, McMaster University.
[12] The Right Honourable Beverly McLachlin, [Former] Chief Justice of Canada, in Adam Dodek, *The Canadian Constitution* (Dundurn Press. 2016), 11.
[13] His excellency the Right Honourable David Johnson, [former] Governor General of Canada, in Dodek *The Canadian Constitution,* 9.
[14]Allan C. Hutchinson, *Democracy and Constitutions: Putting Citizens First* (University of Toronto Press, 2021), 5.
[15] Wil J. Waluchow, *A Common Law Theory of Judicial Review.* (Cambridge University Press, 2007), 19.

constitutions are entrenched – or, made extremely difficult to change – and often require elaborate and super-majoritarian consensus to pass a proposed formal amendment.[16]

I have opened this section with the three quotations above, each of which speaks to one core function of constitutions as they are currently understood: stability, protection, and constraint.

The first quotation, from former Chief Supreme Court Justice Beverley McLaughlin, demonstrates the principle of stability; constitutions are built to last. The entrenched nature of a constitution is intended to promote national stability by limiting both the degree of fundamental change a state can undergo and the speed with which such changes can be made.[17]

Constitutions *should* change; it is impossible to think of all prospective applications of state power in advance. Much like laws in general, a constitution that aims to settle all possible questions and problems in advance will be forced to legislate with such useless generality that it will likely be an ineffective constraint. On the other hand, a constitution that aims to concretely specify all applications will quickly become obsolete if it cannot keep up with changes in social composition, social needs, and technological advancements.

Thus, the ideal constitution strikes a balance between being restraining but not prohibitive; rigid enough to serve as an effective limit on state power, but elastic enough to allow for the changing needs of subjects. Many constitutions therefore include an amendment formula, which outlines the necessary conditions for constitutional change – thus allowing for revision as the needs and values of the community themselves change over time – but aim to do so in a way that promotes stability and graduality.[18] Often a formula will set the bar for constitutional change

[16] Waluchow, *A Common Law Theory*, 41-42
[17] Waluchow, *A Common Law Theory*, 41-42
[18] The amendment procedure for the Canadian Constitution, for instance, states:

higher than the requirements for changes in regular legislation, requiring, for instance, a higher threshold of general approval (i.e. over 50% of the population) on the proposed change, or assent from a certain empowered portion of the population (i.e. the assent of all magistrates), and so on.

The desire for stability comes from reverence for the rule of law. As noted above, constitutions tend to dictate the absolute foundation of the state in question. It will outline how political leaders are empowered, what their powers are, what limits may exist on their powers, and so on. They also serve as a measure of subordinate laws, outlining (for instance) the conditions under which a proposed legal rule is to be recognised as valid, binding law. If these foundations of the social order change too much, or too quickly, it may be difficult for subjects and lawmakers to keep track of what is required of them and act accordingly.

> By this unprincipled facility of changing the state as often, and as much, and in as many ways as there are floating fancies or fashions, the whole chain and continuity of the commonwealth would be broken. No one generation could link with another. Men would become little better than the flies of summer.[19]

This concern about the effects of frequent legal change is recorded as far back as ancient Athens, where it was quipped that "an Athenian who spent three months away from Athens would not recognise the city on his return, because of all the laws passed in the meantime."[20] An additional concern here is that as the number of decisions increase, so too does the risk that decisions become contradictory or inconsistent. With the difficulty of informing citizens of changes in a timely manner and ensuring that those changes are understood, it is very easy for

---

38. (1) An amendment to the Constitution of Canada may be made by proclamation issued by the Governor General Under the Great Seal of Canada where so authorised by

> Resolutions of the Senate and House of Commons; *and*
> Resolutions of the legislative assemblies of at least two-thirds of the provinces that have, in the aggregate, according to the then latest general census, at least fifty per cent of the population of all the provinces."

[19] Edmund Burke, cited in Jeremy Waldron, *Law and Disagreement (*Oxford University Press, 1999), 277.

[20] Wilfried Nippel, *Ancient and Modern Democracy: Two Concepts of Liberty?* (Cambridge University Press, 2016), 66. Notes omitted. The saying is attributed to Plato "the writer of comedies, not the philosopher."

legal requirements to become muddled or confused, risking erosion of the rule of law. As such, a controlled, measured process of legal change – the stability – is highly valuable.

The second quotation in this section, from former Governor General David Johnson, speaks to the fact that constitutions often include a charter or bill of rights which aims to limit legislation around certain protected areas. The purpose of these rights is not simply to serve as some benefit to citizens – they're not mere privilege. Rather, the rights of subjects tend to stand as limits on political authority, and therefore might be thought of as conditions of legitimate government. Political authority is often thought to matter for subjects' practical deliberation; many political theories ground duties of obedience to legitimate political authority to some degree: a leader who infringes on the population's rights may be accused of acting without authority and thus may not be owed obedience.[21]

In the case of constitutional democracies informed by certain political theories (such as liberalism or social contract theory) it has become common for certain rights that are thought necessary to the proper functioning of democratic society – like free expression, or free political association – to be enshrined in the constitution, not simply to serve as checks on state authority but also to withstand interference from the course of regular politics. This prevents, for instance, a group's legal or political rights being affirmed and denied every time political sentiments flop from one political party to another. This can serve to insulate certain minority interests against the whims of a prejudiced or ignorant majority, while also serving the principle of stability.

Many constitutional arrangements include powers for the judiciary to appraise existing law against the constraints of the constitution itself, which allows for an additional route of legal

[21] As I said, the extent to the connection between a leader's authority and a subject's obedience differs across theories. While some theories may ground duties of obedience exclusively on legitimacy, others include additional considerations, and thus an illegitimate authority could still give rise to a binding obligation on the part of subjects even if they are not owed obedience themselves.

recognition should legal rights be infringed, intentionally or otherwise. If, for example, the legislature passes a law that has the unanticipated side effect of infringing on one group's legal right to participation, that group has a route through which to bring a legal case against the legislature and have the conflicting law changed or nullified, to the extent of the conflict, without first needing to pass a majority vote.[22] Since certain political and legal interests may be quite limited in application (e.g., group-based, like indigenous land claims or religious permissions) an inability to muster widespread political and legal support could amount to functional disenfranchisement for those whose interests are not met. Thus, alternate routes like this are valuable as additional or supplementary paths of political and legal access.

The final quotation from Professor Allan C. Hutchinson speaks to what might be the central principle of constitutionalism as such: the exercise of political power must be constrained. Constitutionalism, at its very core, is motivated by the thought that great power carries a risk of great abuse, and that state power must therefore be limited to avoid tyranny.

> If men were angels, no government would be necessary. If angels were to govern men, neither external nor internal controls on government would be necessary. In framing a government which is to be administered by men over men, the great difficulty lies in this: you must first enable the government to control the governed; and in next place oblige it to control itself. A dependence on the people is, no doubt, the primary control on the government; but experience has taught mankind the necessity of auxiliary precautions.[23]

The separation of powers is one method for keeping state power in check and avoiding or slowing the effects of corruption. When power is consolidated, with a single entity responsible for both writing the law and adjudicating the above-mentioned rights-based limits, we're at risk for those rights to become sparse in their content if certain legislative ends are sought, and lack of external oversight means it is unlikely that anyone notices the 'bad behaviour' until it's done.

---

[22] Waluchow, *A Common Law Theory*, 75. It is important to note, perhaps, that support for this principle of protection may vary depending on one's conception of democracy. I will say more on this point later.
[23] Alexander Hamilton, James Madison, and John Jay, *The Federalist Papers: A Collection of Essays* (Coventry House Publishing, 2015), 253. (Letter 51)

By delegating different tasks to independent institutions, there's hopefully less risk of this sort of hollowing-out of rights occurring.[24]

Constitutionalism in its current form has emerged after a long history of badly-behaved authorities and political and social revolution. Forms and flavours of constitutionalism existed as early as ancient Athens, but constitutionalism as we currently know it only really came into its own in the Enlightenment era.[25] Historian Horst Dippel summarises:

> Starting from the medieval idea of a king instituted with *potestas tempurata*, the Glorious Revolution in England in 1688-89 had established the principle of limited monarchy as official constitutional doctrine. The American Revolution, striving to make individual liberty more secure, had transformed this doctrine into the constitutional principle of limited government. For the first time the traditional idea of limiting power had resulted in conferring positive right to the individual, the protection of which he could claim in court against pretensions of the government.[26]

There is nothing inherent in the idea of limiting political power that requires that the power be allocated democratically; indeed, some of the earliest constitutions served to constrain the monarch, and not the *demos*.[27] It's certainly true that some political arrangements may be more compatible with principles of constitutionalism than others. But the conceptual core of constitutionalism is less concerned with where political power comes from, and more with what may or may not be done by whoever happens to wield it.

Constitutionalism and Democracy

The basic principles of constitutionalism hold that constitutions outline the contours and constraints of political power and authority and are (ideally) intended to persist over significant

---

[24] Though, clearly, there's no intrinsic reason why all branches of government can't be equally corrupt or bad faith. This just might ensure that they have different aims in their corruption.

[25] Maurice J. C. Vile, *Constitutionalism and the Separation of Powers* Second Edition (Liberty Fund, Inc. 1998), 23. Notes omitted.

[26] Horst Dippel, "Modern Constitutionalism, an Introduction to a History in Need of Writing," *Tijdschrift voor Rechtsgeschiedenis/Legal History Review* 73, no. 1 (2005): 154-5.

[27] Vile's *Constitutionalism and the Separation of Powers* has an interesting discussion of both the transition from the 'mixed" model of constitution to the separation of powers model, and of how the different 'powers' as we know them (legislation, adjudication, and execution) were picked out as deserving of independent focus. I highly suggest a read.

periods of time to ensure slow, stable, incremental change. These principles seem clearly valuable when compared to absolute monarchy or oligarchy, for instance, where subjects are in danger of the vacillating whims of kings or the negligence of self-serving elites. Overall, then, the picture painted by constitutionalists is bright and enticing, with constitutions serving a pivotal role in ensuring fair and just conditions for legal subjects.

However, whatever the benefits of constitutionalism may be under some political arrangements, the picture seems to change when introduced alongside the principles of democracy. Under democracy, political power is supposed to lie with the people themselves. The point of that common slogan, "by the people, for the people" is to highlight that, in the realm of politics, what the people will is what ought to be done. Representative democracy permits the delegation of this authority to elected officials, but their task remains to enact the will of the people. Whether directly or through elected representatives, the democratic state is intended to put into practice what *citizens* want, and thus political power is thought to belong ultimately to those citizens themselves.

But when a democratic state is constrained by the principles of constitutionalism, citizens and their law are limited by the contents of the constitution in place at the time, which is often one that citizens have inherited rather than chosen for themselves. The result is that current people are thought to be governed by the 'dead hand of the past.'[28] Future generations are subject

[28] Thomas Jefferson is recorded as having expressed concerns about the appropriateness of the 'dead governing the living:

> [N]o society can make a perpetual constitution, or even a perpetual law. The earth belongs always to the living generation. They may manage it then, and what proceeds from it, as they please, during their usufruct. They are masters too of their own persons, and consequently may govern them as they please. But persons and property make the sum of the objects of government. The constitution and the laws of their predecessors extinguished then in their natural course with those who gave them being. This could preserve that being till it ceased to be itself, and no longer. Every constitution then, and every law, naturally expires… If it be enforced longer, it is an act of force, and not of right.

Thomas Jefferson to James Madison, *Volume 15: 27 March 1789 to 30 November 1789* (Princeton University Press, 1958), 392-8

to a constitutional law that is not only not of their making, but also quite difficult to change thanks to the complex amendment formulas. Thus, so long as a protected area (for instance, a right to free speech or mobility) exists as a part of the constitution, no majority however large may legislate against it; formal amendment is required (if possible) before the people may do as they wish. Constitutional constraint – which under any other political arrangement may be nothing but good for the people –apparently ends up working *against* the people's political power in constitutional democracies.

Constitutionalism and Agency

A significant amount of energy has been spent on the issue of whether constitutional constraint is democratically legitimate – a series of concerns known as "the democratic worry." One of the most common rebuttals to this democratic worry insists that a constitution *is* the will of the people, albeit a 'will' of a particular kind. According to this argument, while it is indeed the job of the elected legislature to enact the current, whimsical will of the people, the purpose of the constitution is to outline the *rational* will of that people. According to this rebuttal, in other words, the constitution is a tool of *rational precommitment.* In a constitutional democracy, according to this view, the constitution represents the rational, reflective will of the people – what the people *really* want (or *would* really want) if asked under conditions of cool, calm reflection.

It starts with an acknowledgement that political issues are controversial. Questions of politics have to do with the allocation of social recognition and the distribution of fundamental resources, thereby implicating our fundamental moral beliefs. Where morality is concerned, there is a high likelihood of widespread and vehement disagreement, and (in many cases) a tenuous-at-

best likelihood of compromise and tolerance of difference. Practical disagreement is consequential enough in the day-to-day scene of typical human interaction, but politics isn't strictly a peacetime activity. Political decision-making continues in times of fear and uncertainty, like war and famine; times at which tempers flare and people become wild, vindictive, or desperate. Decisions cannot simply be postponed or dismissed, and we must be wary of making decisions we come to regret.

According to the precommitment view, a constitution and its associated bill or charter of rights aid in defending the rational will of the people by serving as a bastion of our ultimate values and aims. If we have done a good enough job at drafting our constitution, then we will have prepared ourselves for those tense and stressful political moments where calm and deliberate on-the-spot planning may be impossible. We can draft thoughtful deliberative processes that ensure outcomes of a particular moral pedigree, or legally fortify certain political areas we know we will regret entering. Through our constitution we precommit ourselves so that when the time comes for our values to be put to the test, we can come out the other side unscathed. Like the mythical Ulysses, we can pass by the sirens without becoming bewitched and consumed by their song.[29]

I noted previously that the precommitment view is far from some idiosyncratic metaphor of Elster's creation but has significant following. One of the reasons for this, no doubt, is that it can legitimate the practice of constitutional constraint by drawing on the normative force of agency and choice. In western political philosophy, at least, concepts of political legitimacy

[29] The story of Odysseus (or Ulysses) and the sirens has become a common metaphor for the kind of precommitment constitutions are meant to provide. Sirens' songs had the power to make sailors run their ships aground or to jump overboard to their deaths. Odysseus, wishing to hear the sirens sing but live to tell the tale, had the men of his ship tie him securely to the mast of the ship and plug their own ears as they steered within range. Once out of earshot and recovered from his momentary insanity, Odysseus could then be untied. It is this metaphor Elster draws on for both *Ulysses and the Sirens* and *Ulysses Unbound*.

tend[30] to be liberal or 'consent-based': whether a state is politically legitimate in some way turns on whether the state governs with the consent of the people, though the precise flavour of consent differs across different iterations of these views. For one category of views, consent must be explicitly given, with citizens giving direct assent to the state or government; it's not enough that the state makes good decisions or provides ample benefit if the people have not consented to its rule.[31] For the other category of views legitimacy seems to have less to do with whether a person *has* consented, and more about whether they *would* or *ought* to consent given the reasons that apply to them, the values they hold, and the requirements of rationality.[32] Both traditions, which arguably serve as the basis of a significant portion of modern western political and moral thought, rely on the justificatory power of consent; it's not surprising, then, that a practice like constitutional precommitment might also be evaluated according to the same metric. At least some of the concerns about the coercive power of constitutional law and politics seem to disappear if the people have chosen to enact these laws and policies for themselves.

But whether this legitimation is successful is itself controversial. Some are fine with the precommitment view in theory but claim it is often illegitimate in practice, as consent can have no justificatory power where it is not actually given. Thomas Jefferson, for example, argued against the idea of intergenerational constitutions but seemed to allow the possibility that each generation may constitutionally bind itself.[33] If each generation could choose for *themselves* the principles by which their passions are limited, then concerns about democratic legitimacy might

[30] Understand that I am speaking in very general terms. I am not giving a full discussion of political legitimacy. I'm only stating the very commonplace role that 'consent' plays in it.
[31] These are, generally speaking, called 'Lockean' views of political legitimacy.
[32] These are, generally speaking, called "Kantian" views of political legitimacy.
[33] Jefferson famously wrote that "the earth belongs in usufruct to the living," and that "the dead have neither powers nor rights over it." However, he allows that the living be "masters too of their own persons" and may govern themselves as they please, so long as the constitutions they make expire upon their death. *Letter from Thomas Jefferson to James Madison Volume 15: 27 March 1789 to 30 November 1789.* Accessed at: https://jeffersonpapers.princeton.edu/selected-documents/thomas-jefferson-james-madison#notes2b

be satisfied, as consent could more truly be given, and agency more explicitly exercised. But as the critics are right to point out, many of the constitutional provisions by which current generations are bound are *not* ones they have set for themselves.[34] While constitutional amendment is theoretically possible, its difficulty makes it likely that the constitution in effect has often been inherited from a bygone generation. For this reason, constitutional precommitment, insofar as its bonds span beyond those who bound themselves, remains undemocratic.

Advocates of the precommitment view have two general types of responses intended to satisfy these objections. The first form of argument – which I call the *democratic precondition* argument – tries to pre-empt the democratic worry by arguing that constitutional precommitment is a necessary precondition for, and therefore cannot be incompatible with, democratic self-government. The basic core of this argument is that rather than *limiting* the democratic process, constitutional constraints *constitute* the very conditions necessary for democracy to take hold at all.

> A "basic law" is introduced as the necessary and sufficient condition for the democratic process itself, not for its results: democracy cannot define democracy. The relationship between democracy as the source of legitimation and a constitutionalism that does not need democratic legitimation poses no paradox, however. For constitutive rules that first *make a democracy possible* cannot *constrain* democratic practice in the manner of externally imposed norms… enabling conditions should not be confused with constraining conditions.[35]

Of greatest importance here is the *denial* of any conflict, rather than a recognition of an *existent but justified* conflict. To be clear, constitutional constraints are thought to be justified.

---

[34] My favourite version of this argument, for its sheer attitude, comes from Waldron, who has said that anyone who thinks that constitutional precommitment is like the self-binding from the story of Ulysses (Odysseus) and the Sirens "is an idiot." (Waldron, *Law and Disagreement*, 268) For an equally sassy response, see Wil Waluchow, "Constitutions as Living Trees: An Idiot Defends" (2005).

[35] Jürgen Habermas, "Constitutional Democracy: A Paradoxical Union of Contradictory Principles?" *Political Theory* 29, no. 6 (2001): 770. Emphasis original.

> These [justifications] range from traditional liberal fears of 'the tyranny of the majority' that suggest a conflict between individual rights and democracy… through to proceduralist or more republican accounts that stress [certain constraints] as either facilitating or intrinsic to the democratic process themselves.[36]

As a necessary precondition for democracy, constitutions don't *need* democratic justification, since they cannot be undemocratic; the justifications noted above are simply the icing on top.

The view is common. Ferrajoli argues that fundamental rights are necessarily a part of democracy, and the role of the constitution (and judicial review) is to protect them.[37] Barak has also said "democracy has its own internal morality" which must be "insulated from the power of the majority."[38] Dworkin's justification of judicial review rests on the assumption that it restores a democratic imbalance, insofar as the overturned law does not satisfy essential democratic preconditions.[39] Holmes argues that constitutions "like the rules of grammar" aren't just rules that handcuff or restrain participants; instead, they "make participation possible, by setting out a matrix of interaction in which particular contributions can take their place and 'register'." [40] Without a constitution, democracy simply wouldn't be possible, since there'd be neither protections for one's right to political participation, nor an established process through which this right could be exercised. Constitutions are a necessary step in the democratic process, and thus operate outside judgments of democratic illegitimacy.

I suspect that the democratic precondition argument may be quite compelling under certain conceptions of democracy, but democracy is a highly (perhaps *essentially*) contested

---

[36] Richard Bellamy and Dario Castiglione, "Constitutionalism and Democracy- Political Theory and the American Constitution," *British Journal of Political Science* 27, no. 4 (1997): 598.
[37] Cited in Victor Ferreres Comella *Constitutional Courts and Democratic Values: A European Perspective*. (Yale University Press, 2009), 87.
[38] Cited in Comella, *Constitutional Courts and Democratic Values*, 87.
[39] Cited in Comella, *Constitutional Courts and Democratic Values*, 87.
[40] Quoting Waldron, *Law and Disagreement,* 276. Original in Stephen Holmes "Precommitment and the Paradox of Democracy." In *Constitutionalism and Democracy*, edited by Jon Elster and Rune Slagstad (Cambridge: Cambridge University Press, 1988), 227.

concept; there is significant disagreement about what democracy is and requires, which results in different views about what the necessary preconditions of democracy are going to be. Those who take democracy to be more substantive in form and moral content – such as what we might call liberal or republican conceptions of democracy[41] – may be more likely to agree with the precommitment view, as the constraining power of a constitution may be just what is needed to protect citizens from interference or domination, as the case may be. To those for whom democracy is something more strictly majoritarian, however, such constraint may itself amount to interference and domination. What one group thinks is the bare minimum required for the democratic process is, for another group, an impediment to the democratic process; and while we could easily dismiss one or the other as being simply 'mistaken' about 'what democracy really is,' it is unclear (at least to me) who may more rightfully claim to have the 'correct' meaning of democracy in hand.[42]

As such, the defence that constitutional precommitment is a precondition for democracy seems to beg the question. Some are willing to concede that even strictly majoritarian conceptions of democracy need *some* sort of prior arrangement to function, but they think that the full picture of constitutional precommitment as it commonly exists (with bills of rights and judicial review) take normative constraints too far.[43] Neither the advocates nor the critics of

[41] The different conceptions of democracy, it seems to me, vary according to their understanding of the concept of liberty. By 'liberal' conceptions of democracy, I have in mind those conceptions that operate on "liberty as non-interference" or "negative liberty," and "liberty as freedom to" or "positive liberty." Republican conceptions, such as those associated with Phillip Pettit, operate according to "liberty as non-domination." There are certainly many more conceptions, but the point is only to highlight some different kinds. I am not giving a taxonomy of democracy or liberty.

[42] A full discussion of the question about the 'real' conception of democracy is not the focus of this dissertation. I certainly have views about what claims of conceptual correctness entail and require, but this is not the place. As a note, what will be 'correct' depends, at least in part, on the kind of conceptual project one is doing, which will certainly digress into an argument about what the 'correct' conceptual project is.

[43] James Allan, who argues for a 'thin' conception of democracy, makes such a point. See Allan, "Thin Beats Fat Once Again – Conceptions of Democracy" *Law and Philosophy 25* (2006): 533-559.

constitutional precommitment take themselves to be working from an *anti*-democratic point of view; but they *do* have different takes on what the democratic point of view looks like.[44]

The second type of defence for the precommitment view – which I call the 'metaphysical' argument – is given in response to arguments like that from Jefferson, which tolerate the precommitment view in theory but note that most constitutional constraints have not been chosen by current generations. Generally, the metaphysical argument aims to defend the idea of a "people in a diachronic sense"[45] – that is, the idea of an atemporal, collective identity – by whom and for whom the precommitments have been chosen and are thus binding. On this view, the relevant 'self' who brings about the commitments by which current citizens are bound is a *collective* self consisting of all who 'belong' to the broader political category or identity. Ronald Dworkin describes a metaphysical argument in the following statement:

> If I am a genuine member of a political community, its act is in some pertinent sense my act, even when I argued and voted against it, just as the victory or defeat of a team of which I am a member is my victory or defeat even if my own individual contribution made no difference either way. On no other assumption can we intelligibly think that as members of a flourishing democracy we are governing ourselves.[46]

This view could be understood as a version of social contract theory, or like Rousseau's 'general will.' The constitutional commitments are not the desires of any one generation, but instead the will of the general people, where "general people" includes all citizens, past and present (and future). While no individual living person has set the constitutional commitments for themselves, it's nevertheless the case that the collective to which they belong has historically bound itself. So, for example, the Canadian constitution isn't representing the will of *current*

---

[44] For another discussion of these conflicting accounts of democracy, see Bellamy and Castiglione "Constitutionalism and Democracy."
[45] Cristina Foroni Consani "Pré-Compromisso Constitucional e Autonomia Coletiva: Uma Conciliação Possível?" *Revista de Estudos Constitucionais, Hermenêutica e Teoria Do Direito*, 7, no. 3, (2015): 238.
[46] Ronald Dworkin. "The Moral Reading and the Majoritarian Premise" In *Deliberative Democracy and Human Rights* ed. Ronald C. Slye (Yale University Press, 2008), 101.

Canadians, but instead represents the will of "Canadians" as a collective entity over time. And, they say, since the constitution *does* represent the will of the collective people, the democratic worry is allayed.

Of the two defences of constitutional precommitment, the metaphysical argument seems the weaker, as full support of it commits us to some strange conclusions about moral agency. Jeremy Waldron has argued precisely this point, stating that buying into this argument "strains the credibility of the elements of the self, agency, and autonomy involved in the precommitment idea." Like Jefferson, Waldron thinks that the commitments made by one generation could, perhaps, be thought of as commitments for *that* generation, but certainly not for those who follow.[47] Elster also demonstrates this line of thinking in his later works, arguing that the collective cannot be understood in the same way as the individual; while *as individuals* we are thought to have one part of our selves responsible for our well-being and long-term planning, there is no part of the collective which has any inherent claim to represent the general interest.[48] As such, the success of the metaphysical argument rests on a very strange understanding of people as agents, and is a pill some aren't willing to swallow.

For what it's worth, I think constitutional precommitment is – *and is meant to be* – democratically illegitimate. My full opinion on the matter would look something like the democratic precondition argument, albeit with a different punchline. I *do* think constitutional precommitment is necessary for the rule of law in democratic societies, regardless of the form of democracy. But where I disagree with the democratic precondition argument is in the conclusion that the *fact* of this necessity somehow inoculates the practice against claims of democratic

[47] Waldron, *Law and Disagreement*, 270-2.
[48] Jon Elster, *Ulysses Unbound,* 92-94; 167-168. Society, says Elster, has "neither an ego nor an id."

illegitimacy. Constitutional precommitment, to me, *is* undemocratic – that's quite the point of a constitution.

My interest in the democratic worry and associated arguments isn't primarily focused on the worry or arguments themselves. Instead, I find the discussion interesting because, despite the widespread disagreement on just about every other part of the dialectic – political legitimacy, the concept of liberty, the concept of democracy, and so on – there *is* agreement between critics and advocates alike that agency and choice play a central role in the theory.

Now – obviously – I'm not mystified about why agency might matter for debates about democratic legitimacy. It doesn't take a huge leap of the mind to see that, if agency is important for liberty, and liberty for the concept of democracy, thus agency is important for the concept of democracy. I'm not saying we know or agree on what agency is or why it matters, nor whether we know if or why it matters for liberty; I'm only pointing out that a connection between democracy and agency, even *prima facie*, seems intuitively clear.

But my interest isn't in the connection between agency and democracy. Rather, I am interested in the perceived connection between agency and commitment. The upshot of the democratic worry debate seems to me to be that agency is *as* central to constitutional precommitment as it is to democratic theory, such that if constitutional precommitments aren't chosen, then they can't really be commitments at all. But I think the error in the precommitment view, as its currently understood, is how it overemphasises agency. If agency is *not* conceptually necessary for commitment, as advocates and critics both seem to assume, then the question of whether constitutional precommitment is democratically legitimate is only materially significant for one's own reflective equilibrium.

It should come as no surprise that I happen to think agency is *not* as central to commitment as is often thought. Whether certain constitutional commitments are democratic or not does not matter for whether the people are nevertheless committed; chosen or not, constitutional commitments are the present generation's commitments. This idea is certainly compatible with our *wanting* democratic legitimacy or wanting agency to have a central role in the commitments people have. Jefferson and Waldron might be right to argue that it's unfair or unbalanced for one generation's chosen constraints to bind future generations. But this, I think, is quite different from concluding that one cannot have commitments that one has not chosen for oneself.

The presumed truth of the agent-dependent condition is not unique to constitutionalism – it's simply the most interesting application of the error that I've seen. The problem, it seems to me, is a conflation present in the ordinary use of commitment itself. The concept of commitment is often used in place of associated concepts like enthusiasm, investment, or value. We might talk about "our commitments" when we really mean to talk about the values we think are worth protecting, or the relationships we think are worth developing. Through the natural laziness of ordinary conversation, we have collapsed the spaces between the *commitments* we have and *our reasons* for having them, letting statements about the things we value serve as statements about our commitments themselves. Commitment, then, becomes synonymous with the choices we make.

But to me, our commitments are not our choices themselves, but the parameters within which our choices are both made and made possible. Of course, what we value – our relationships, morals, investments, etc. – affects the structure of those parameters to a certain degree; this is why it would be incorrect to completely ignore the role of agency when discussing

commitment. But we will struggle to explain how these important ideas – like agency, valuing and choosing – work if we assume commitment to be nothing more than what is assumed by the agent-dependent condition. Agency plays a role in commitment, but commitment is not simply a consequence of agency.

This discussion is only a brief glimpse into the view I will put forward in later chapters. As should be clear by now, I do not think commitments only exist where we've chosen them, and thus I think that the binding-others problem is not, in fact, a problem. Before we get there though, I wish to consider another misconception about commitment, again through the vehicle of constitutional precommitment. Never mind that constitutional precommitments may only bind others; even more damning is that they may not bind anyone at all.

Chapter 3: The (Non)-Constraint of Commitment

We could think of the two problems of constitutional precommitment – the non-bindingness problem and the binding-others problem – as expressing polarised concerns about the constrainingness of constitutional commitment. The binding-others problem from the previous chapter takes constitutions to be so constraining that they create a condition whereby the living are bound by the dead. The core of the binding-others problem, then, might be thought of as a kind of constitutional claustrophobia; constitutions have *so* much power to control and constrain that the very thought causes distress. The core of this next problem, however, is that constitutions may lack the power to bind anyone or anything at all.

There are two slightly different senses of this problem, which I'm calling the non-binding problem. The first sense is discussed by James Madison in Letter 48 of the *Federalist Papers* and expresses a doubt in the power of a constitution as anything more than a "parchment barrier."

> Will it be sufficient to mark, with precision, the boundaries of these departments, in the constitution of the government, and to trust to these parchment barriers against the encroaching spirit of power? This is the security which appears to have been principally relied on by the compilers of most of the American constitutions. But experience assures us, that the efficacy of the provision has been greatly overrated; and that some more adequate defense is indispensably necessary for the more feeble, against the more powerful, members of the government.[49]

The problem in this first sense is simply that a constitution – as a series of words written on a document, or some set of abstract values or ideas – can neither declare nor enforce its own will. If there is no one to consider or defend the constitution, the constitution will not consider or defend itself.

The second sense of the non-bindingness problem is that which comes from Elster's conception of precommitment in *Ulysses and the Sirens*. While the first sense, as expressed by

[49] Madison et al, *The Federalist Papers*, 48.

Madison, takes aim at the constrainingness of constitutions generally, Elster's sense of this objection seems limited to his particular understanding of precommitment; constitutions might constrain in some way, but that way isn't properly considered "precommitment."

In this chapter, I'll devote greater focus in discussing the second sense of the non-bindingness problem, since it is more clearly connected to the agent-dependent conception of commitment. Given how much it seems to me that the agent-dependent condition arose out of an illicit extension of Elster's work on indirect (or imperfect) rationality, I want to dedicate a good amount of time to explaining what he took his initial project in *Ulysses and the Sirens* to be and explain where I think he went wrong in his 'sequel' *Ulysses Unbound*. Where my diagnosis of the binding-others problem in the previous chapter was an overemphasis on the role of agency as the source of commitment, my diagnosis of the non-bindingness problem is a mistaken belief about the *kind* of constraint commitment is.

Elster 1979: *Ulysses and the Sirens*

While *Ulysses and the Sirens: Studies in Rationality and Irrationality* is a collection of essays on rational decision-making, the greatest portion of Elster's first text focuses on (what he calls) *imperfect* rationality. The title of this text refers to a story from Homer's epic *The Odyssey* pertaining to the enchanting song of the sirens and a sea captain who wished to hear it. The captain, Ulysses[50] was aware of the danger[51] of hearing the sirens' sing, and thus had his crew tie him firmly to the mast of the ship so he could survive the sirens' song. The crew plugged their

[50] Aware of the fact that this is the Roman name of the Greek Odysseus, I will nevertheless maintain the use of Ulysses for consistency's sake.

[51] The different iterations of the story vary on exactly what the song did to listeners. In some cases, hearing the song drove listeners mad, causing them to leap overboard. In other cases, it is merely that the song distracted sailors into crashing into rocks, inadvertently sinking their own ship. Either way, the point is that it is a challenge to hear the sirens sing and live to tell the tale.

own ears, thereby avoiding the same trance while remaining free to steer the ship beyond the sirens' domain.

> Ulysses was not fully rational, for a rational creature would not have to resort to this device: nor was he simply the passive and irrational vehicle for his changing wants and desires, for he was capable of achieving by indirect means the same end as the rational persons could have realised in a direct manner. His predicament – being weak and knowing it – points to the need for a theory of imperfect rationality.[52]

In Ulysses' case, there is a sense in which rational thought is impossible once transfixed by the sirens' song, given some sort of magic or witchcraft is at play. But in ordinary people the problem is ordinary weakness of will. Presumably, if human beings were fully and perfectly rational, then weakness of will would not be an impediment to rational thought that we needed to overcome. But weakness of will is our reality.

Given the title and nature of the text, and the portion of the text dedicated to this discussion of imperfect rationality, it's clear that Elster's intentions with this work are to explain cases in which these methods of indirect rationality are indeed *methods of rationality* – rather than states of irrationality or passivity – which have been "all but neglected by philosophers and social scientists."[53] Elster's theme may be rational decision-making, but he's keenly aware that perfect rationality is, at best, an incomplete picture.

As such, Elster puts forward a skeletal theory of one method of indirect or imperfect rationality. Elster's general understanding of the Ulysses method, which he calls self-binding or precommitment, is that the method "is a privileged way of resolving the problem of weakness of will; the main technique for achieving rationality by indirect means."[54] He does not have a

---

[52] Jon Elster, *Ulysses and the Sirens,* 36. Emphasis added.
[53] Elster, *Ulysses and the Sirens*, 36. Elster does note the 'pathbreaking' work done by R.H. Strotz in consumer behaviour and George Ainslie on experimental psychology, both of whom he credits with having "laid the empirical and conceptual foundations" for all later work. His project is "a first step toward a synthesis" and an effort to "broaden the empirical and conceptual base itself."
[54] Elster, *Ulysses and the Sirens*, 37.

definition proper, but suggests that there are five criteria on which any proper[55] definition of precommitment must bear:

1. To bind oneself is to carry out a certain decision at time t1 in order to increase the probability that one will carry out another decision at time t2.[56]
2. If the act at the earlier time has the effect of inducing a change in the set of options that will be available at the later time, then this does not count as binding oneself if the new feasible set includes the old one.[57]
3. The effect of carrying out the decision at t1 must be to set up some causal process in the external world.[58]
4. The resistance against carrying out the decision at t1 must be smaller than the resistance that would have opposed the carrying out of the decision at t2 had the decision at t1 not intervened.[59]
5. The act of binding oneself must be an act of commission, not of omission.[60]

The first requirement highlights the *intentionality* of Elsterian precommitment. The increased probability of the decision at t2 *must* be the motive of the bond made at t1 for the bond to be one of precommitment. With this first condition, Elster means to exclude cases where a desired outcome has emerged as an unintended side effect of another action or an unpredictable but welcomed effect of the engaged action.[61] Elster's example of the smoker who desires to quit smoking is a useful case for clarifying this distinction. Self-binding would require the smoker, for example, to refuse to buy more cigarettes once he runs out (and even, perhaps, to dispose of the ones he has already). Intentionally running out of cigarettes would obviously increase the likelihood that he will not smoke at a later time, and this state of not smoking is directly connected to the decision made at time t1 not to buy more cigarettes. This would be contrasted

[55] In this discussion, references to "real" or "proper" precommitment or self-binding are all evaluations made according to Elster's view. I mean nothing more, here, than to say that Elster would consider it precommitment (or, if it's not "real" precommitment, then to say he wouldn't).
[56] Elster, *Ulysses and the Sirens*, 39.
[57] Elster, *Ulysses and the Sirens*, 42.
[58] Elster, *Ulysses and the Sirens,* 42.
[59] Elster, *Ulysses and the Sirens*, 44.
[60] Elster, *Ulysses and the Sirens*, 46.
[61] Elster, *Ulysses and the Sirens*, 40.

with the smoker who forgets her pack of cigarettes at home. Not having her cigarettes right in hand might bring about the desired state (not smoking), but she cannot be said to have precommitted to quitting smoking. The difference between the two cases is the exercise of agency: while the former case includes intentional deprivation, the latter case of forgetting one's cigarettes does not. As such, though both cases may bring about the state of non-smoking, the latter is merely passive, not 'imperfectly rational.'

The second requirement is what separates precommitment from other forms of self-binding, such as delaying gratification, as Elster does not think that precommitment amounts to a sacrifice of present goods "in order to make more goods available later". If the goal of the bond made at time t1 is to preclude action *y* at time t2, then a *proper* bond should preclude *y* as an option at time t2. Elster seems to require some irreversibility in this understanding of what it takes to be properly committed. Thus, if *y* is still a feasible option at t2, the 'commitment' is, in essence, reversible and thus not commitment. A person who struggles with abusing their prescription medication has not precommitted themselves against this outcome if they have merely left their medication in another room and 'put it out of their mind,' as retrieving the medication is always easily within the realm of possibility. Using a controlled medication dispenser or asking the pharmacist to dispense a limited amount of medication at a time represents a greater impediment to abuse and, as such, a more 'properly' Elsterian bond.

The third requirement is necessary to preclude what Elster calls "decisions to decide"[62] and is what he thinks sets the Ulysses method apart from other methods of imperfect or indirect rationality. While Elster does not think this sort of 'steeling oneself' decision-making process is always doomed to fail, he *does* think that processes of this sort tend to have very little impact. It

[62] Elster, *Ulysses and the Sirens*, 43.

ought to be clear why: as Elster notes, the very reason for engaging in this precommitment in the first place is out of recognition of a weak will and/or strong external pressure. Especially in cases of repeated failure to, for example, quit smoking or use prescribed medications responsibly, "present failure will predict future failure."[63] In this case it would be irrational, in a known condition of weakness, to depend on the strength of my will to get me through the temptation. Thus, Elster thinks it necessary for self-binding that we deposit our will into some "external structure," even if only temporarily.[64] 'Structure' here is somewhat undefined, but in conjunction with the previous criteria, seems just to refer to a deliberate modification of the environment to make the desired change in behaviour more likely. As seen from the explication of the second criterion, if *y* is to be avoided at time t2 yet remains a feasible option, then one has not precommitted oneself.

The fourth requirement is a response to humanity's "general resistance to walking uphill, and our preference for downhill strolls."[65] The chain smoker who wishes to quit might find quitting outright to be too difficult, and so may embark on a journey to strengthen her willpower. But if going to the gym to instill discipline or taking up a hobby for distraction are going to require the same willpower needed to quit smoking, these will be poor strategies for following through on her commitment.[66] Instead, the smoker might commit herself to an indirect method, one that resembles a "one step back, two steps forward" compromise. This might include steadily decreasing the number of cigarettes she may have a day, knowing that while she is still smoking, she will reach a point where the habit is sufficiently subsumed by her willpower that she can achieve the final state of 'not smoking.'

---

[63] Elster, *Ulysses and the Sirens*, 44.
[64] Elster, *Ulysses and the Sirens*, 43.
[65] Elster, *Ulysses and the Sirens*, 44.
[66] Elster, *Ulysses and the Sirens*, 46.

The fifth and final requirement holds that "the act of binding oneself must be an act of commission, not omission."[67] Elster does not explain this criterion much beyond noting the difficulty of precisely distinguishing 'omission' from 'commission' and noting that while it is possible for one to bind oneself by omission, he thinks such a case is not precommitment proper. His justification for including this criteria rests on his noticing that "the fact that someone prefers not to leave a given state is not evidence that he would have freely entered that state from all of the states that are open to him."[68] As his example, Elster considers the story told of moral education:

> It does not seem adequate to me to say with Gerald Dworkin that 'an important moral limitation on the exercise of parental power…is provided by the notion of the child eventually coming to see the correctness of his parent's intervention' for virtually any educational scheme could be extended so as to produce its own justification ex post facto. Rather I would say that the mark of a successful education is that the child comes to see that no such justification is possible, but that the parents nevertheless had to make some (unjustified) choice.[69]

There is an insight in this example that I will return to later. But for now, it serves only as Elster's example of why acts of omission - here understood to be the adult child *not* going against what their parents taught them – should be separated from the kinds of acts that Elster thinks are appropriately precommitment. Precommitment must require some deliberate choice not to do X, rather than the mere absence of a choice to do X.

In discussing existing alternatives to the Ulysses method, Elster considers several arguments about what imperfect rationality might look like on a more collective level.[70] One set of arguments he considers is in the realm of politics, where he mentions the "problem of democracy."

---

[67] Elster, *Ulysses and the Sirens*, 46.
[68] Elster, *Ulysses and the Sirens*, 47.
[69] Elster, *Ulysses and the Sirens*, 47. Notes omitted.
[70] Elster, *Ulysses and the Sirens*, 85.

> A direct democracy, either in the sense that all citizens vote on all issues instead of electing representatives, or in the sense that representatives can be recalled at any time – will tend towards zig-zag policies and constant reevaluation of past plans; it will be incontinent, vacillating and inefficient.[71]

Elster considers some of the solutions[72] that have been given to deal with the problem of democracy and discusses whether he thinks they would pass muster as *true* (i.e. Elsterian) precommitment. It is during this discussion where, in a passage no more than a paragraph long, Elster writes what he will later denounce as "a mistake."

> At another level the system of periodic elections could be interpreted in the same perspective… In this interpretation periodic elections are the electorate's method of binding itself and of protecting itself against its own impulsiveness…We also noted Lindbeck's suggestion that the politicians in turn could bind themselves so as to avoid this temptation, by randomly spaced elections. This would provide a new answer to the problem of who shall guard the guardians, and indeed the randomly spaced elections could be the electorate's device for simultaneously binding itself and the politicians.[73]

For the record, Elster does continue analysing the problem of democracy beyond just this passage. He highlights the problems of constituent assemblies and the fact that their unique and privileged characters *as* the constituent assembly are held only by historical accident. In considering the paradox of democracy – that each generation wants to be free to bind its successors while not being bound by its predecessors[74] – Elster concludes that "later generations have no obligation to feel bound by their predecessors, but neither do they have any legitimate right to bind their successors."[75]

As should be clear from the above discussion, Elster's focus is very much centered on the rational agency of the individual, and how that agency is not only compatible with but *served* by methods of indirect rationality like the Ulysses method. His discussion of the electorate binding

---

[71] Elster, *Ulysses and the Sirens*, 88.
[72] One author he considers is M.I. Finley, who considers the practice of ostracism and 'graphe paranomon' as precommitment devices for ensuring political stability.
[73] Elster, *Ulysses and the Sirens*, 90. Lindbeck's proposal is meant to counteract the politician's temptation to "lump all unpopular measures at the beginning of the electoral period and popular ones at the end."
[74] Elster, *Ulysses and the Sirens*, 93.
[75] Elster, *Ulysses and the Sirens*, 95.

themselves was clearly, simply, a clarification of the kind of thing that *could* be considered precommitment compared to other examples he'd considered through the course of his analysis. His conclusion about what rights later generations have over themselves and their successors is highly reminiscent of his earlier statement about childhood moral education; while perhaps no justification can be given for what's done in the name of stability, the problem of democracy and its vacillating, topsy-turvy mechanisms require *something* (however unjustified) be done.

Despite what I think was his careful effort to cleanly delineate his project, Elster's conclusions in *Ulysses and the Sirens* have been given much broader application than warranted. What seems to me to have been one example of a mechanism that could provide indirect or imperfect rationality was taken to be an exact, definitive statement about what Precommitment is.

Waldron, for instance, takes Elster's third criterion at exact face value.

> Jon Elster has suggested that a decision at t1 counts as a way of binding oneself vis-à-vis some decisions at t2 only if '[t]he effect of carrying out the decision at t1 [is] to set up some causal process in the external world'… In the political case, what counts as an external structure?...Does this include or exclude strategies like the drinker giving his car keys to his friend? The friend's possession of the car keys is not really a causal mechanism ensuring or increasing the probability that the drinker will not drive home at midnight. Instead, it operates by virtue of the friend's undertaking at t1 not to give back the car keys at t2 together of course with the friend's willingness at t2 to actually honor that undertaking. My point is not that the friend may prove unreliable; for so may a causal mechanism. It is rather that the precommitment device operates via the friend's judgment and decision, and to that extent its operation at t2 is not entirely under the drinker's ex ante control at t1.[76]

Waldron's problem here might be one of the following. First, he may be expressing concern over the fact that the friend to whom I've given my keys may have a different interpretation of relevant concepts. Problems like this are problems of communication and interpretation as such. My friend, for instance, may have a certain behavioural expectation of

[76] Waldron, *Law and Disagreement*, 260-2.

drunkenness which I do not fit, on the basis of which they return my keys to me when I objectively should not have them (or deprive me of my keys when I am objectively, fully and completely sober). Given the theme of *Law and Disagreement* (where this discussion occurs) is, aptly, law and disagreement, this seems a charitable and likely interpretation of the issue.

If the issue is exactly the matter of disagreement, and relatedly whether the friend "makes the ethically appropriate decision" I can only say that I think Waldron expects the concept of precommitment (and commitment) to do much more than it does and can be reasonably expected to do. Waldron has in mind a scenario where my infant child becomes ill during a party where everybody is either drunk or unlicensed, on the basis of which he argues that causal mechanisms are too deontological or rigid. In giving my friend my keys and instructing them not to let me drive home drunk, it is not clear where I'm importing the condition that they act as Judge Hercules,[77] assessing the circumstances to make the best moral or ethical decision. Rather, most people, as Waldron points out, would quite simply refuse to return the keys – which is, after all, what I asked them to do. Should my child need medical attention, then my inebriation and lack of access to my keys is an unfortunate circumstance (which I perhaps should have considered in advance). My friend and I may perhaps renegotiate the circumstance of the original precommitment, given this novel situation; maybe we'll consider the moral wrong of driving drunk versus driving unlicensed. But moral perfection is neither a necessary condition nor necessary aim of precommitment.[78] The fact that some precommitment device might result in an outcome that is anything other than morally ideal does not detract from the practice of

[77] Reference to Ronald Dworkin's example of Judge Hercules, who is posited as the perfect moral reasoner. See Dworkin, *Law's Empire* (Harvard University Press, 1986.)

[78] This is likely going to result in philosophical disagreement, much in the same way people disagree about whether evil promises are 'real' promises. I acknowledge the opportunity for discussion but will not be engaging in discussion here.

precommitment or from the friend's role as a precommitment 'device;' it was a bad play, but it was a play nonetheless.

This is where the second interpretation of Waldron's concern comes in, and while the first interpretation is more in line with the themes of *Law and Disagreement*, the second interpretation, I think, is in line with how discussions tend to treat the concept of (pre)commitment itself. The second interpretation of Waldon's problem with the friend as causal mechanism is that the *fact* that my friend's judgment is not within my control is itself significant; that, since this is supposed to be *my* commitment, anything less than full control detracts from its status as my commitment. This, recall, would align with the agent-dependent condition, which, recall, I think is mistaken. If we assume that commitments are only constraints that we put on ourselves, then my lack of control over my friend's judgment upsets this direct link.

Whether we see the friend operating as a causal mechanism or merely as a source of external judgment, the *point* of the precommitment is that the events of t2 are not within the agent's control. Waldron seems to think that at t2 – at the point of commitment – all parts of the prior precommitment ought to remain within the agent's control to some extent. He very specifically states that his problem "is not that the friend may prove unreliable; for so may a causal mechanism." But if the problem with the friend's judgment is *not* that they may ultimately be a poor causal mechanism, the remaining issue seems to be only that the friend is not within my control. But this concern is odd to me; for, if the point of precommitment is to counter act weakness of will, then it seems that my lack of control in this moment is precisely the point.

Elster, 2000: *Ulysses Unbound*

By the time we get to his second work in this area, *Ulysses Unbound*, Elster has had a change of heart, no doubt owing to the objections raised by Waldron and others against

constitutional precommitment. If it is correct to think that the five conditions are present in any real commitment, then it seems that constitutional precommitment faces some problems.

First, since it's a feature of precommitment, as Elster understands it, that individuals bind *themselves*, constitutional self-binding seems to be a misnomer; after all, it is not individuals who have bound themselves to their constitution, but the historical constituent assembly, founding fathers, or the framers. This is, recall, the binding-others objection, which was the focus of the previous chapter. These commitments are not made by the current government, and yet limit the powers of the current government – this is, recall, the core of the democratic worry. This would mean, it seems, that rather than being a commitment to bind one's own behaviour, constitutions serve only to bind the behaviour of others.[79]

The problem for Elster here is that societies are not "individuals writ large."[80] In an individual, weakness of will arises because the self is – in an important sense – divided; one part of me really wants to cut my hair and dye it green, but the other part knows I'll regret it soon after. Elster argues that even though the self is not unitary, there is nevertheless a part that is 'in charge' and "can engage in long-term planning to restrict the myopic or impulsive action tendencies of the other parts."[81] The self, otherwise put, has a vertical division. But society is not like this, as there is no group who has an inherent claim to represent the general interest more than any other. The divisions in society are horizontal, not vertical: "Society has neither an ego nor an id."[82]

---

[79] Elster, *Ulysses Unbound*, 92.
[80] Elster, *Ulysses Unbound*, 167.
[81] Elster, *Ulysses Unbound*, 168.
[82] Elster, *Ulysses Unbound*, 168. I'm not making any claims here about what the 'self' is 'really' like. These are Elster's statements only.

We saw in the previous chapter that the binding-others problem rests on the (I think mistaken) belief that commitments – to be *real* commitments – must be chosen by the committed party. While Elster adds the additional ego/id argument about the disanalogy between individual and collective self-binding, it nevertheless still depends on the hidden assumption that commitments must be chosen; the only implication of the disanalogy argument is to explain why the constraint is justified in the individual case and not the collective.

This makes sense for Elster; his whole purpose in *Ulysses and the Sirens* was to put forward an initial framework for indirect or imperfect *individual* rationality. His goal was not to put forward a theory of commitment or precommitment, but rather to consider the conditions under which these ideas are compatible with rational autonomy. As such, it makes sense that for Elster a more direct line from the individual's agency to the commitment is required. The *mistake*, I think, was that Elster (et al) seems to have forgotten that while he may have put forward a theory of imperfect rationality, he certainly did not provide a conceptual analysis of commitment or precommitment as such. In other words, precommitment as a method of imperfect rationality is but a small part of the whole picture of commitment. Any arguments about constitutional precommitment which rest on Elster's understanding of precommitment, then, are going to be limited by Elster's intentionally narrow project.

The binding-others problem is, to me, hardly the greatest hurdle for constitutionalism to clear. Potentially more troublesome is the claim that *constitutions may not bind at all*. If constitutional constraints can be changed (or simply ignored) by those who don't like them, then how can constitutions fulfill their constraining function at all?

Constitutional amendment is a good thing; a constitution which does not permit change may end up preventing sensible, optimal behaviours. We must tread lightly to avoid suffocating

citizens or legal actors, tempting them to ignore the constitution or to protest its authority. We thus implement "safety valve clauses" or amendment formulas. If we can formally amend the constitution, then our constitution will be able to keep up with changes in the world and in the populace.

The challenge, Elster notes, is that "if the framers try to prevent the constitution from becoming a suicide pact, it may lose its efficacy as a suicide prevention device."[83] The problem with constitutional amendment is that while it makes change difficult, there is yet a sense in which everything that can be done can be *undone*. This is, it seems, incompatible with Elster's criterion that the action to be avoided must not exist in the set of feasible options at t2. If, through amendment, that which was removed from the set of actions at t1 may be reintroduced into the set of options at t2, then the second criterion hasn't been met.

The problem of formal amendment doesn't even consider the fact that constitutions cannot enforce themselves. It's one thing for a group to decide they shall change their constitution, and thus participate in the pre-ordained processes for doing so; it's quite another thing for that same group to simply declare they are not bound by the existing constitution and decide to pave their own way. In such a case, a constitution is unlikely to represent much more than "a parchment barrier."[84]

> The individual who wants to bind himself can entrust his will to external institutions or forces, outside his control, that prevent him from changing his mind. But there is nothing external to society, barring the case of precommitment through international institutions with powers of enforcement such as the International Monetary fund or the World Bank. And even these cannot make it impossible to act against a precommitment, only make it more costly to do so.[85]
>
> And just like that, Elster reverses his original position.

---

[83] Elster, *Ulysses Unbound*, 174.
[84] Hamilton, *The Federalist Papers,* No.48.
[85] Elster, *Ulysses Unbound*, 94-5.

> In *Ulysses and the Sirens* I came close to claiming both that constitutions are precommitment devices (in the intentional sense), and that societies ought to bind themselves by constitutional precommitment devices. As I have been saying in various places earlier [in this text], these claims are eminently contestable.[86]

In *Ulysses and the Sirens*, Elster muses about constitutional precommitment as a kind of collective equivalent to his theory of imperfect rationality. He ultimately abandons this possibility – even as a partially valid one – since, he thinks, he did not "fully understand the extent of the disanalogy between individual and collective self-binding… For one thing, constitutions may bind others rather than being acts of self-binding. For another, constitutions may not have the power to bind at all."[87]

Do Constitutions Bind?

It seemed to me that the binding-others problem from the previous chapter came about because of an overestimation of the role of agency in commitment – a mistaken belief about commitment that I have called the agent-dependent condition. The non-bindingness problem, however, arises out of what I think is yet another mistaken belief about commitment and the kind of constraint that commitment represents. Constitutions *do* bind; we're simply mistaken about what the bond of a commitment looks like.

Statements about the constraining function of constitutions suggest to me that the *kind* of constraint that commitment creates is interpreted far too literally. Perhaps our imaginations have been corrupted by the frequent allusion to Ulysses, or perhaps a general anxiousness with coercion and constraint narrow our sights to the most central instances of each. But I find it especially surprising that Elster retreats as he does, rather than double down on some of his original statements from *Ulysses and the Sirens*. From my interpretation of his earlier statements, he was very much on the right track.

---

[86] Elster, *Ulysses Unbound*, 167.
[87] Elster, *Ulysses Unbound*, 92.

Amidst his conclusions in *Ulysses and the Sirens,* Elster reiterates the numerous methods of imperfect rationality, which he lists as:

1. Manipulation of the feasible set
    a. Restricting the set of physically possible actions
    b. Changing the reward structure by public side bets
2. Manipulation of Character
    a. Strengthening will-power
    b. Changing the preference structure
3. Manipulation of Information
    a. Changing the belief system
    b. Avoiding exposure to certain signals[88]

Elster expands

> Within the set of actions (1) that change the feasible set I distinguish between, say, (1a) the strategy of going for a long walk in the mountains so as to make cigarettes physically unavailable and (1b) the strategy of telling your friends that you will stop smoking so as to change the reward system. The (1b) latter strategy also induces a change in the feasible set, because the option 'Continue to smoke without any sarcastic comments' now becomes unavailable. Within (2) character-modifying actions I distinguish between (2a) the general strategy of strengthening the will-power and (2b) the more particular strategy of modifying some particular desire; the first permits you to climb higher uphill slopes while the second lowers the height of the slope that is to be climbed. Within the set of actions (3) that modify the information upon which further decision are taken I distinguish between (3a) the radical strategy of inducing new factual beliefs (including a strategy for inducing forgetfulness about the induction) and (3b) the moderate strategy of screening yourself from certain signals or cues in the environment.[89]

You'll notice, in particular, that the method of changing the reward system (1b) is remarkably metaphysical. While much of the discussion of constitutional precommitment and precommitment as such seem to centralise method (1a), Elster himself recognises the limited

---

[88] Elster, *Ulysses and the Sirens*, 103. Elster mentioned a fourth strategy which fails as a method of precommitment proper, as it fails to satisfy his third criterion. He includes it for completion, as it is a given response to the problem of weakness of will. The fourth strategy is

4. Manipulation through a rearrangement of inner space
    a. Using private side bets
    b. Using consistent planning

[89] Elster, *Ulysses and the Sirens*, 103-4. I include the long quote to state precisely Elster's understanding of how these strategies differ.

usefulness of this (and other) methods, stating that he thinks "the most important of these strategies are probably methods (1b) and (2b)."

> In most actual cases strategy (1a) is unfeasible or involves sacrifices too heavy, and would thus fail to satisfy criterion (iv) of II.2. The same, I think, holds for method (2a), which would often involve a form of overkill whose feasibility might do away with the problem itself... I submit that in most everyday cases the most efficient strategy involves a combination of methods (1b) and (2b). Realising that through a series of actions I can achieve a hexis from which the desired actions will flow naturally ('sans violence, sans art, sans argument'), and that each of these actions is outside the immediate reach of my willpower, I may precommit myself to them by changing the reward system.[90]

To me, it seems that discussions of constitutional precommitment expect that all commitments are or must be of (1a). Again, one of the major impediments to the coherence of constitutional precommitment is that everything which can be constitutionally done can be *undone*, and thus nothing is ever really removed from the realm of possibility.[91] Or so it seems.

The insight from the long quotes above that I wish to highlight is contained in Elster's statement about method (1b). There, again, he says

> Within the set of actions (1) that change the feasible set I distinguish between, say, (1a) the strategy of going for a long walk in the mountains so as to make cigarettes physically unavailable and (1b) the strategy of telling your friends that you will stop smoking so as to change the reward system. The (1b) latter strategy also induces a change in the feasible set, because the option 'Continue to smoke without any sarcastic comments' now becomes unavailable.[92]

Notice that in method (1b), it's not *smoking* that is impossible. Elster's understanding of the option which have been 'removed from the feasible set of options at t2" is not one's ability to smoke. Instead, what is impossible is *smoking without sarcastic comments.* This is a highly intangible kind of constraint and seems wildly different from the kind of constraint wherein smoking *itself* is made physically impossible as in the example for method (1a).

---

[90] Elster, *Ulysses and the Sirens*, 104-105. As a reminder, criterion iv is "The resistance against carrying out the decision at t1 must be smaller than the resistance that would have opposed the carrying out of the decision at t2 had the decision at t1 not intervened."
[91] Waldron, *Law and Disagreement*, 260.
[92] Elster, *Ulysses and the Sirens*, 103. Emphasis added.

This is important. The non-bindingness problem often (if not always) portrays precommitment as requiring that constitutional commitments resemble method (1a), where a particular action itself – for instance, the infringement of a political right – becomes *physically impossible*. The threshold of constitutional constraint that such arguments presuppose is so rigid that it seems as though we are to imagine a legislator or judge's hand being – quite literally – tied. Obviously, as both Waldron and Elster rightly point out, no such measure of constraint exists: constitutions are parchments barriers, and they cannot enforce themselves when disrespected or overlooked.

Given their absence in *Ulysses Unbound*, I think Elster himself may have forgotten some of his conclusions about the relative usefulness of some of the different precommitment methods from *Ulysses and the Sirens*. His statements about the best methods being those like (1b) changing the reward structure by public side bets or (2b) changing the preference structure are noticeably absent in his discussion of constitutional precommitment. Instead, much like Waldron and others who critique the precommitment view, Elster seems to have come to understand the constraint of commitment more rigidly and literally as (1a) restricting the set of physically possible actions, despite his earlier claims that such methods were overkill, or required sacrifices too heavy to be feasible.

Again, if it's really the case that commitment requires a physical impossibility, then constitutional precommitment is indeed impossible: constitutions can never bind us in the way Ulysses was bound to the mast. But I think there is a great value to considering what commitment as a far more metaphysical or intangible kind of constraint might look like, not just as a general theory of commitment, but also for the practical application of constitutional precommitment in particular. In other words, if we change our understanding of the constraint of

commitment to something less literal and less physical, then we change the picture of what constitutional precommitment could look like: it would no longer be necessary that constitutions bind our hands. Not only has Elster already considered intangible constraints, like public side bets, to count as precommitment, but he said himself that he thinks they are the most efficient methods of self-binding. I think Elster was already on the right track.

Chapter 4: Agent-Dependent Conceptions of Commitment

The relationship between commitment and *precommitment* is temporal. *Precommitment* is used to pick out the period of time *before* one is committed in a particular situation: I arrange to leave my cigarettes at home *before* I leave the house, so that I am unable to smoke when I have cravings later; or I give my keys to my friend *before* I start drinking so that I am unable to drive home drunk later. As such, *pre*commitment is a useful concept for considering isolated instances of commitment.

But though it is coherent in this limited instance to consider some state of 'before being committed' and 'after being committed,' this language describes a very narrow, very singular instance of commitment. Precommitment is thus a bit of a misnomer; it's simply commitment by another name, an artificially described condition picked out from the broader commitment matrix in which it is found and from whence its power comes. In actuality, there is no time at which one is *not* committed. There is no time before or after commitment, unless we count the non-experience of before life and after death. Commitment and precommitment are, ultimately, one and the same. The mechanism of precommitment can only work because it takes advantage of the constraints that are already operative around us; if those constraints did not exist, then precommitment itself would not be possible. As such, a significant amount of work can be done for precommitment (and by extension, constitutional precommitment) if we invest time in the concept of commitment itself.

One issue, it seems to me, is that the way we speak of commitment has affected how we understand it in such a way that commitment seems to amount to little more than 'personally significant choice.' If we recognise any aspect of constraint associated with commitment, then we do so in the manner described by Elster and Waldron; there's some superego who is

preventing our id from acting out, but *we* – the whole person – are the ones responsible for the constraint.

I don't want to deny that there are examples, like those considered by Elster, of our ability to leverage commitment to serve rational or long-term interests. But I *do* want to deny that commitment is *nothing more* than this. The role of agency in determining whether and why one is committed is vastly overstated, and I think this limits how we can think about downstream theories, like those of constitutional precommitment. What is needed is a discussion of commitment that is not limited to the conditions under which one can be 'rightly' considered an agent, as was the condition of Elster's project.

Thus, in this chapter I consider commitment itself, both in everyday thought and in academic work. One of the issues is that there is relatively little material for both; while we might frequently speak of commitment in our everyday speech, very little conceptual work has been done on the matter. Sam Shpall's work on distinguishing moral and rational commitment is, thus far, the only focused conceptual work on commitment that I have found; all other discussions of commitment in academia presuppose a 'folk' understanding and treat the concept only instrumentally (such as "How to determine romantic commitment" or "How to create committed employees"). Shpall, much to my excitement, is interested in the concept itself.

The order of this chapter is as follows. I begin with a brief look at the 'folk' conception of commitment, and I rely on common 'commitment' statements or conditions to draw out certain intuitions about how a 'typical' person tends to understand commitment. This method is crude and far from scientific, but I do think my conclusions tend to hold. The folk conception of commitment is incomplete, but I don't think it is completely ungrounded from the more

academic conceptions of commitment (and precommitment) from Shpall and Elster. Thus, while I'll be criticizing it, the folk conception is worth mentioning.

I will then turn to Shpall's "Moral and Rational Commitment," in which Shpall argues for the distinction between *rational commitment* and *moral commitment*, while insisting that they do indeed belong to a common concept of *commitment (*which is contrasted with the concept of *requirement)*. Shpall's vocabulary and chosen examples are significantly useful, not only in seeing how he understands commitment, but also in explaining where I disagree. While I do disagree with Shpall on some major, central points, there are parts that I like and, given he has written the only direct analysis of commitment that I have found at present, it would seem disrespectful not to include it.

In the end of the chapter, I want to draw out some issues I see, both with the folk conception and with the 'academic conception' (used here to refer broadly to Shpall and, to some extent, Elster) and suggest a different understanding. There are two issues that I centre on. First, the conceptions of commitment that I discuss all overemphasise the role of agency as constitutive of commitment: in other words, they all operate according to the agent-dependent condition. Second, owing to this overemphasis on agency, the constraining power and nature of commitment has been misrepresented, resulting in (what I think are) redundant or *ad hoc* arguments about 1) the variability of commitments from person to person, and 2) the difference between commitment and 'different' normative standards, like requirement. On my picture, commitment is, in a sense, the end-all-and-be-all of normativity: concepts like requirement, which are typically treated as distinct from the concept of commitment, are little more than refined sub-sets or bundles of commitment. And if it really is the case that those requirements

(which we tend to hold in common) are in fact commitments, then there is significantly less variability of commitment from person to person than Shpall assumes.

Commitment: The Folk Conception

In ordinary language, the word 'commitment' tends to evoke associated ideas like devotion or allegiance, or common instances of commitment, like marriage or promises. We may say things like "she is a committed musician" to describe something about a person's character, such as their reliability or persistence. In these cases, being committed appears to be something said of a person in their behaviour toward that to which they are committed. By saying "I am a committed musician," we may be saying something about how frequently I practice, how long I have been playing, or perhaps something about my enthusiasm and how I value being a musician over other obligations I may have.

But despite this common association, it seems to me that we also *don't* think commitment is nothing more than good-feeling or enthusiasm. In saying "he is a committed parent," we might be saying something about how he ranks his obligations, with those parental obligations taking precedence over work and social obligations. But we don't, I think, conclude that someone fails to be a committed father because he sometimes wants time away from his children, or that someone who sometimes dislikes their job necessarily fails to be a committed employee. The folk conception of commitment recognises and, I think, can accommodate the possibility that one might 'feel' committed to a thing, and yet sometimes experience negative emotions or thoughts associated with that thing; being committed seems to refer to the fact that one follows through, even when one might wish to be free to do otherwise.

This idea of following through despite wistful or wishful thinking certainly seems to capture some of the constraining aspects of commitment. Commonly given examples of commitment, like marriage, work, or parenting, are all recognised conditions of sacrifice; one is

constrained in a very particular way. Conversely, we also don't typically think of those who frequently change jobs, hobbies, or partners as being 'committed' to those jobs, hobbies, or partners, even if they were fully and completely enthusiastic for the duration of that job, hobby, or partnership. These ideas of value and constraint, then, are both going to be somewhat important for understanding commitment, as what it means to be committed doesn't just boil down to good feelings about that to which we're committed.

One understanding of the relation between value and constraint in commitment has to do with value serving as the *source* of constraint. If I'm a committed parent, then I am constrained by the obligations that go along with parenting and will frequently have to subordinate some other activities or desires in order to follow through on those parental obligations. But the reason for subordinating those other desires is that *I* place high value on being a committed parent; there are associated constraints on my behaviour, but they are ultimately constraints that I have placed on myself because of the value I place on parenting.

This more active understanding of the constraint – being limited because I've chosen to be – would certainly fit with the agent-dependent condition, as well as with what Elster (and, as we'll see, Shpall) has to say about precommitment. But we don't only speak of commitment in this active sense. We might also use 'committed' in a more passive sense to express our being backed into a corner or having reached a point of no return. Frequently in philosophy, for example, we speak of how someone is "committed to a conclusion" – not in the sense that they argue for it devoutly, but in the sense that if they wish to maintain their existing beliefs, then they must accept certain implications, whether they want to or not.

To me, this more passive use of commitment is illuminating. The choice to believe X results in a condition where, because of the relationship between X and Y, we would say one is

committed to Y. But one hasn't *chosen* Y, at least not the same way one has *chosen* X. While my belief in X might be fully and completely within my control, it seems to me that there is a sense in which my relationship to Y stands somewhat apart from it, insofar as I do not have control over the relationship between X and Y. This passive sense is something that Shpall picks up on in his work, though he doesn't refer to it as such.

Both Shpall and Elster agree that the folk concept of commitment is misleading or inaccurate. Ordinary use tends to conflate 'commitment' with 'enthusiasm' or 'investment,' which hides the core of what precommitment and commitment are intended to capture: a certain independence from these very feelings of enthusiasm or investment.

Both think that, while commitment *is* 'agent-dependent' in one sense, there is yet an important sense in which the concept of commitment is much more constraining than the folk understanding tends to convey. I will be explaining Shpall's view on this in due time. For Elster, this is more of a deduced conclusion. Elster's project was of course on *pre*-commitment; his point was to discuss what one would be required to do at t1 in order to be able to claim rational agency (albeit, indirectly) at time t2. Where t1 is the instance of precommitment – the instance at which one puts in place the causal mechanisms that will alter one's behaviour – time t2 is the instance of commitment. It is at t2 that one is intended to be constrained. Thus, though there isn't, it seems, a point at which Elster directly declares that he shares this view, it seems like there is good evidence to support the claim that he does nevertheless recognise this particular kind of constraining role that commitment plays.

While I agree with the belief that commitment is much weightier than is thought under the standard 'folk' conception, I think claims that commitment is 'agent-dependent' (or, to put it differently, that one can *only* be committed by oneself) are misleading at best and outright wrong

at worst. While no doubt the things we value or the choices we've previously made are part of the picture, I think that any account of commitment that centres *only* on an agent and their choices is lacking. While academic work on commitment may be less guilty of conflating commitment and enthusiasm, there is still an assumption (or a skewed perspective, perhaps) that the self is the only (or ultimate) source of 'real' commitment - an assumption that I think is mistaken.

Sam Shpall: "Moral and Rational Commitment"

For Shpall, commitment is a normative relation that can exist in two forms: rational commitment and moral commitment.[93] While distinguishing between these types of commitment, Shpall's aim here is also to argue that these are indeed *two types of commitment*; two forms that ought to be thought united under the same conceptual heading. I will begin with rational commitments.

For Shpall, "to be rationally committed to [believing] A is to be such that you must be irrational if you fail to [believe] A, assuming no changes in your other attitudes."[94] In essence, rational commitments are epistemic in the sense that they refer to what we ought to believe, given the beliefs we already hold. Thus, in the example Shpall puts forward,[95] if Adam believes everything in the bible to be true, and Adam also believes that the bible says the world was created in six days, then Adam is committed to the belief that the world was created in six days, or he would be irrational.[96]

---

[93] Sam Shpall, "Moral and Rational Commitment." *Philosophy and Phenomenological Research* 88, no. 1 (2014): 146.
[94] Shpall, "Moral and Rational Commitment," 148.
[95] Shpall, "Moral and Rational Commitment," 147
[96] We can, of course, expand on these options: Adam could either denounce his belief in the bible, or qualify his belief. He could deny that his belief is in the bible's 'literal' truth, and instead believe it only in the same way that one might think it true that they really were "taking the hobbits to Isengard."

There are two features common to both rational and moral commitment: first, for each, commitment is a normative relation. Put nicely by Shpall:

> An agent's actual attitudes constitute the ground of his rational commitment – that is, what makes it the case that he has the commitments that he does – but not the commitments themselves. The commitments themselves are normative, in the sense that they put genuine pressure on the committed agent to form the attitude to which he's committed; and this pressure obtains independently of how he thinks about it.[97]

The first thing of note here is that Shpall attributes the *ground* of commitment – what makes it the case that one has been committed – to one's agency. Under Shpall's view, one is only committed when one in some way takes on the commitment for oneself, such as by holding or expressing a belief, or (in the case of moral commitments) through some communicative act. But while Shpall is clearly centralising the role of agency for the question of how one comes to be committed, he does not think agency extends to the question of what one's commitments are. I may 'choose' to hold a certain belief, such as believing in the truth of the Bible, the consequence of which is the series of subsequent or derivative conclusions that follow from that first belief. But despite having chosen to hold that first belief, I do not get to choose what conclusions I am therefore committed to; that, for Shpall, is the normative force of commitment.

This first feature captures what I previously called that 'passive' sense of commitment. But even while recognising that rational commitment has normative force outside of the individual's attitudes, Shpall thinks this normative force is limited. For Shpall, being rationally *committed* to a belief is not (and ought not be thought) the same as being rationally *required* to believe it. Requirement, here, is used to refer to conditions or norms that capture some sense in which an individual "falls short of some important kind of rational (or moral) ideal."[98] This constitutes Shpall's second feature of rational commitment.[99] Adam is not rationally *required* to

---

[97] Shpall, "Moral and Rational Commitment," 149. Emphasis original.
[98] Shpall, "Wide and Narrow Scope." *Philosophical Studies* 163, No. 3 (2013), 717.
[99] Shpall, "Moral and Rational Commitment," 149.

believe that the world was created in six days in any objective sense.[100] To claim otherwise would amount to one irrational belief sanctioning another in an "illegitimate sort of bootstrapping" with irrational belief all the way down. Thus, what Adam might be rationally committed to is, to Shpall, "intuitively" distinct from what he is rationally required to believe. I disagree with this claim and will return to it later.

Thus, the two features of rational commitment are that 1) they are minimally normative, but 2) are not requirements. These two features also hold for moral commitment.[101] Just as I can be rationally committed to a belief independent of my actual attitudes about the belief itself, I can be morally committed to, for example, keeping my promise even without the intent or desire to do so. But, same as above, moral commitment is not the same as moral requirement; I may be morally *committed* to complete an unjust promise (since I *have* promised) but it'd be a mistake to say that I'm morally *required* to complete it.[102]

The difference between moral and rational commitment, then, has nothing to do with these features; these are what unify them as two different types of commitment. Instead, Shpall thinks that the differences pertain to what each type of commitment takes as their object, and how commitments of each type come about.[103] Rational commitments take one's own mental states as their object and come about through one's rational activities: I acquire certain rational commitments by "forming beliefs that stand in certain relations," and are therefore grounded in my actual mental states.[104] Moral commitments, on the other hand, come about through our

---

[100] To some, he may even be rationally required *not* to believe this.
[101] Shpall, "Moral and Rational Commitment," 150.
[102] This, obviously, will be contingent on whether one thinks we can really make 'immoral' promises. If the theory of promising precludes evil or impossible promises, then it seems no such distinction would arise. It seems contextually true that Shpall must accept that we may make promises that are evil, impossible, or never intended to be kept.
[103] Shpall, "Moral and Rational Commitment," 151.
[104] Shpall, "Moral and Rational Commitment," 151-152.

dealings with other people and thus, for Shpall, require us to perform some communicative action.[105] Shpall limits his discussion (whether intentionally or accidentally) to the communicative act of promising; thus, it seems, one only comes to have a moral commitment through the act of making a promise.

Shpall highlights two core differences between moral and rational commitments which are of greatest significance here. First, while he thinks both rational and moral commitments may be undone, the method of doing so differs across types. Shpall writes,

> In the [case of moral commitments] securing release involves contacting the promisee – again, performing a communicative action, and obtaining some kinds of permission from someone else. In the [case of rational commitment] being released from a commitment simply requires revising or discarding some of your attitudes.[106]

Second, moral and rational commitments differ in what Shpall calls "their normative natures." Recall, in the example above, Adam is rationally committed to the belief that the world was created in six days or else he is irrational. However, Shpall thinks this normative structure does not carry over to moral commitments.

> The analogous claim would be: to be morally committed to ϕ-ing is to be such that you must be immoral if you fail to ϕ, assuming no changes in the grounds of your commitment. But a moment ago we gave a counter example: Obama could be morally committed to spending money on green jobs even if failing to do so would not be immoral.[107]

The example he speaks of is his illustration of moral commitment. In it, as noted, President Obama commits in late January to spending ten billion dollars on green jobs. But in early February, there is a devastating natural disaster in a densely populated area. It would seem, Shpall writes, that "Obama is morally required to divert the ten billion dollars from green job creation to humanitarian assistance," thereby breaking his moral commitment. But surely, Shpall

[105] Shpall, "Moral and Rational Commitment," 151.
[106] Shpall, "Moral and Rational Commitment," 152.
[107] Shpall, "Moral and Rational Commitment," 152.

concludes, the arrangement is such that Obama is not immoral for this breach in his prior commitment. Thus, he concludes, the analogous claim must not hold.

In summary, Shpall enumerates the differences between moral and rational commitments with a series of facts:

1. Acquisition Fact: Rational commitments are unilaterally acquired; moral commitments are not.
2. Escape Fact: Rational commitments are unilaterally escaped; moral commitments are not.
3. Violation and Conflict Fact: Violation and conflict of rational commitment is always avoidable, whereas violation and conflict of moral commitment is not always avoidable.[108]

Furthermore, for the purpose of his paper, Shpall also assumes the truth of what he calls the permissibility principle:

4. Permissibility Principle: Violation and conflict of commitments is impermissible if not properly excused.

And these facts, in conjunction with the permissibility principle, lead us to the "interesting conclusion" about the "divergent normative natures" of moral and rational commitment:

5. *The Normative Difference*: We can always rationally fulfill, or escape, all of our rational commitment; but it is not always the case that we can morally fulfill, or escape, all of our moral commitments.

Thus, for Shpall

> The nature of rational commitment guarantees that we are definitely irrational when we violate a rational commitment. But the nature of moral commitment does not allow for such a guarantee. Sometimes the violation of a moral commitment entails no immorality – as we have already observed above, in the case of Obama's diverting the green jobs funds to humanitarian assistance.[109]

[108] Shpall, "Moral and Rational Commitment," 163.
[109] Shpall, "Moral and Rational Commitment," 163.

Discussion

While I like a lot of what Shpall has to say, I take issue with his view on the normative difference between moral and rational commitments, as well as his distinction between commitment and requirement.[110]

First, I'm not convinced that the normative difference between moral and rational commitment holds, as Shpall seems to rely on incommensurable examples to draw this conclusion. In the example of Adam and Creation, Shpall notes that Adam must be irrational if he does not also believe that the world was created in six days, absent any change to his other attitudes. However, Adam's avoids irrationality by changing one or all of his beliefs (as noted above); release from rational commitment simply requires revision or rejection of those attitudes that grounded the original commitment. If he no longer believes everything in the bible is true or that the bible says that the world was created in six days, then he is no longer rationally committed to the belief, and is thus no longer irrational if he fails to take that belief on board. Shpall acknowledges this option, making the point only that unless the grounds of his existing views change, Adam is rationally committed to the belief that the world was created in six days. With all this, I agree.

Adam is rationally committed to believing that the world was created in six days *because* he believes in the truth of the bible (*and* claims himself to be rational). Thus, the nature of his rational matrix is thus: Adam can *break* this commitment by maintaining his existing belief in the bible while rejecting the conclusion that the world was created in six days; he could *complete* this commitment by accepting the conclusion and taking on board the belief that the world was created in six days; and finally, he can *undo* the commitment by rejecting the truth of the bible,

[110] I also take issue with the "permissibility principle" but that discussion will be left for my discussion of legitimacy in Chapter 6.

or by changing one or all of the existing beliefs which initially committed him to the conclusion that the world was created in six days.

I am 'irrational' if I fail to believe that to which I am rationally committed, *absent any changes to the grounds of my commitment.* Shpall thinks this does *not* hold for moral commitment, hence the 'normative difference'. But this seems mistaken, since it ignores what I think are important parallels between moral action and rational belief. Obama, like Adam, has three options: he can (1) *complete* his commitment by following through on paying out the funds for green jobs, as promised; (2) *break* his commitment by failing to follow through despite the initial conditions of the commitment remaining the same; or (3) *undo* the commitment by showing that the grounds of the initial commitment have changed.

Shpall concludes that, because we don't think Obama has done something immoral by failing to fund those green jobs, there must be a difference between the normativity of moral and rational commitments. But this only seems plausible because he has built shifting grounds into the Obama example, and thus the examples aren't commensurable. The true analogue for the Adam case would not be Obama reneging on a promise because of a sudden, pressing, and substantial natural disaster – that disaster would arguably represent a significant change to the conditions under which the original promise for funding was made. Instead, the parallel would be if Obama did not fulfill his pledge in the *absence* of such a disaster, in which case Obama likely would be criticized and declared immoral. In the stated example, the sudden natural disaster has shifted the grounds of the original promise. This is why it appears like Obama did not act immorally despite not completing his moral commitment– his commitment wasn't *broken* but was *undone*.

Thus, Shpall's argument for the apparent normative difference between moral and rational commitment rests on a tilted comparison between a rational commitment which has been *broken* and a moral commitment which has been *undone*. When we consider comparable examples of broken moral and rational commitments, there doesn't appear to be any normative difference: in commensurable examples, I think Obama *would* be judged to have acted immorally.

The second issue I take with Shpall's view is that he presupposes a difference between commitment and requirement which I don't think is warranted. For the record, while I strongly disagree with the distinction between commitment and requirement, I recognise the intuition that Shpall is drawing on here, insofar as I understand the desire to purify concepts that are used to impose strong normative pressure; *requirement*, when used in this way, is one such concept. But while I can appreciate the concern about 'illegitimate bootstrapping' of irrational beliefs, there nevertheless seems to me to be no real distinction between commitment and requirement.

> Consider the example of Adam and Creation again. Of the example, Shpall writes:
>
> According to the dominant view of writers on rationality (the 'wide-scope conception of rational requirement), it is simply not true that Adam is rationally required to believe that the world was created in six days… Really Adam is rationally required to either form the belief that the world was created in six days, or to revise his belief that everything the Bible says is true… This claim is supported by our linguistic practice, as we may appropriately say things like
>
> (7) Adam is committed to believing some crazy things; but what he should do is stop believing that everything that the Bible says is literally true.
>
> Whereas it would sound at least somewhat awkward to claim that
>
> (8) Adam is committed to either believing that the world was created in six days, or to revising his beliefs about the Bible.
>
> On the face of it, then, rational commitment is distinct from rational requirement.[111]

[111] Shpall, "Moral and Rational Commitment," 149-150.

Maybe I've simply missed something, but (8) does not seem odd or awkward to me. In both cases, the hidden premise involved in the judgment is the claim that Adam wishes to be (or perhaps, *ought* to wish to be) rational. If we already believe, as Shpall does, that commitments bear normative pressure regardless of one's desires, then it's not clear to me why presenting both disjuncts of (8) as making up part of Adam's commitment matrix should seem intuitively different from stating what he is 'required' to do. In fact, in this particular, focused example, it seems like the rules of logic might *require* Adam to believe the world was created in six days *just as much* as they might require him *not* to believe in the Bible. Shpall clearly rejects such a claim as "simply not true,"[112] but fails to consider where the force of this rejection is coming from. Should one wish to be rational, then there are indeed several conditions that must be met, some of which likely result in Adam's need to abandon his belief in the Bible altogether. But Adam needn't be rational; claims of what he *should* rationally do don't matter to Adam unless he claims he is (or wishes to be) rational.[113]

This, of course, is a direct point of disagreement with Shpall. Shpall thinks commitments are agent-dependent, but requirements are not. He writes, again:

> By claiming that commitments are agent-dependent I mean that, in order to come into existence, a commitment must be grounded in an activity of the agent who is to become committed. This is not meant to imply anything ambitious about free agency. The idea is just that since both moral and rational commitments depend, in a broad sense, on the activities of the agent who comes to stand in the commitment relation, it makes sense that different agents are committed to different things. Another person – or, more generally, the world itself – does not have the power to commit you to actions, intentions, or beliefs.[114]

---

[112] Shpall, "Moral and Rational Commitment," 149. He references only the existing literature on "wide-scope views about rationality." While these models provide an explanation of his rejection, I'd be curious to know his thoughts on what makes the tenets of these theories (or any other theories) anything but formalized commitment matrices.
[113] This is an issue subject to major philosophical disagreement. I'm not getting into the debate around reason internalism or externalism, as it's just not the time or place. I recognise that this is not a universal belief.
[114] Shpall, "Moral and Rational Commitment," 154. Emphasis added.

But I disagree. In the case of rational commitment, 'rationality' refers to a particular structure of thought such that to *be rational* is to reason according to the established, pre-existing rules. Absent these rules, Adam is perhaps perfectly capable of holding contradictory beliefs. It's only on the introduction of logic as the rule and measure of 'proper' thought that Adam's existing beliefs are declared irrational. But Adam is not the source of logical rules, neither in his attitude nor his actions. Those are imposed – by others, by analytic philosophers or logicians, or by the world – regardless of whether Adam has done anything to put himself within their jurisdiction. He could indeed *claim* to be rational in his beliefs, which would be a more overt demonstration that he means to be assessed according to the rules of logic. But even still, the normative force of rationality (and thus, rational commitment) comes not simply from the beliefs one holds, but from the metric against which those beliefs are assessed. And those metrics are not agent-dependent, nor is the tendency for one's beliefs to be judged accordingly. As such, it seems like agent-dependence can only get us partway to understanding how commitments come to be held.

Of course, Shpall's response is to likely say that the rules of logic (or moral laws) are *requirements*, not commitments, and therefore they *are* imposed by the world. But again, this overemphasises the role that agency has in determining whether one is committed. It seems to me a more accurate description that both commitments and requirements are imposed by the world – since, it seems to me, requirements are but a refined or specific subcategory of commitment. The word 'requirement' may be used to distinguish those normative pressures that we feel *justified* in imposing on others, but requirements are simply commitment under another name.

The problem, I think, is that Shpall (like Elster), is once again relying on observations from a limited areas of investigation to draw general conclusions about the nature of the category as a whole. In the case of requirement, Shpall is using a metric of assessment in the theory of rationality to draw conclusions about commitment as such. To me, this has the same effect as studying the rules of chess to draw conclusions about the nature of games. While we may certainly glean some useful and important information about games in general by studying the rules of chess, it would be a mistake to think that full knowledge of chess provides full knowledge of games as such; but it would be equally mistaken, I think, to think that the differences between chess and other games mean that chess is not a game, but must be some kind of independent category of rule-based behaviour.

Requirement is a commitment game. When we say something is required of someone, we're making a series of claims about the nature of the commitments we think they have and the substantive values attached to the commitment matrix in which (we think) they exist. We might be making a claim about how some commitment carries with it a certain kind of social attitude: saying something is *required* might mean that we think coercion is justified to ensure compliance. It might betray something about the generality of the commitment, such that what is "required" refers to a more broadly-held expectation – something that might be thought to be universally true, for instance. Or, saying something is *required* may mean nothing more than "You might not like it, but you have to do it." But while the word "requirement" may be used in these cases, that does not mean it is not commitment; requirement might be conceptually *refined,* but it is not conceptually *distinct*.

Thinking of requirements as commitments is an option only open to me because I do not centralise agency in the same way that Shpall and Elster do. If we remove the assumption that

commitments must be chosen by oneself to be binding, then it isn't clear what remaining value there in in distinguishing between commitment and requirement, at least as far as rationality is concerned.

The case for moral commitment is perhaps more difficult to make, partly owing to Shpall's incredibly narrow treatment of moral commitment as promises.[115] On first glance, it seems plausibly correct to think that if moral commitments are just promises, then moral commitments are indeed agent-dependent, as seemingly no one can promise on your behalf. But even this doesn't seem correct. We've likely all been the child whose parent promised our attendance at some very boring event, and subordinate workers are frequently committed by their superiors to fulfill the terms of a contract, or to follow a set of imposed regulations. Perhaps Shpall might say in response that such examples are not promises. But surely the conclusion to draw from such examples is not that these are not commitments, but rather that promises is too limited a conception of moral commitment.

Shpall justified the agent-dependent requirement by claiming that it alone "explains why agents' commitments vary widely."[116] Thus, without agent-dependence as a feature of commitment in general, we'd be unable to explain how we can all have such different commitments. But this is begging the question. Granted, if moral commitment really is so limited as to be merely 'the promises we make,' then it's indeed the case that our moral commitments vary widely. But to maintain this position is to ignore how much we have in common. We are all forced to participating in a similar legal, economic, or political system by dint of proximity. These carry forth more specific commitments, such as needing to work for income, needing to

[115] To be fair, he does not explicitly state that *only* promises are moral commitments, but instead takes promising to be the "paradigmatic case" of moral commitment. Nevertheless, promising ends up being the only case of moral commitment he considers here.
[116] Shpall, "Moral and Rational Commitment," 154.

pay rent, bills, and taxes, and needing to know certain languages or methods of communication. Certainly, some have the means to exist outside of these constraints and, perhaps, may not need employment. But nevertheless, it is arguably the nature of social living that those around us share many (if not most) of the same commitments.

But while we share many commitments, Shpall is right to note that something needs to explain the variation that does exist. But I don't think absolute agent-dependence – in the sense that one is only committed if one has chosen for oneself – is needed to perform this task, as acknowledgment of our different places in the world – both literal and metaphorical – seems sufficient.

Think back to the intro philosophy class where one first comes across a particularly frustrating thought experiment, such as the trolley problem. The frustration of such a moral dilemma – whether to save the five at the expense of the one or let the five die – is that the agent deliberating has done nothing to bring the scene about, and yet in many ways is inextricably linked to its outcome. Had they not been walking in the area that day, the scene would nevertheless have still played out (perhaps with some other poor sucker at the helm). But once someone is aware of the scene, their moral landscape has fundamentally changed. The nature of such an event is that one has no choice but to act, even if that action is *inaction*, merely because they stumbled upon an already unfolding scene. If one does not act, one will be judged for failing to act; if one does act, then one's actions will be judged as either right or wrong. This *is* the world imposing moral commitment. *What* the agent chooses is up to them, but *having to choose* has been thrust upon them.

The argument could perhaps be made that in both cases these moral commitments come about because of something the agent has done: in the trolley case, the agent came upon the

scene; in Obama's case, he ran for the office of president and its associated duties. However, even if this argument has teeth, it is yet a far cry from Shpall's claim that moral commitments come about through communicative acts.[117] Obama may never have communicated any intention to provide humanitarian aid in the event of a natural disaster; and I have never stated any intent to stumble upon a live-action trolley problem. If by 'communicative action' Shpall has in mind something less literal than an utterance or statement of intent, then it's unclear which of our actions become communicative. Do I communicate the intention of telling another the time merely by wearing a watch? Do I communicate an intention of driving another to an airport merely by having a car?

The main problem, I think, is that Shpall insists on distinguishing commitment from requirement, but his effort in cleaving these ideas apart is unrefined. One possibility is that the agent-dependent condition of commitment is so pervasive that it simply escapes the imagination to consider commitment in any other way. Another possibility is that a desire for conceptual purity drives us to insulate certain kinds of normative pressures – like requirements – from being tainted by irrationality or immorality. But an alternate, less agent-centric conception of commitment might help us satisfy these conditions while forcing us to confront the false sense of security such conceptual purity generates.

I agree with Shpall that there are different types of commitment, and that they differ with respect to their creation, their normative pressure, and the conditions under which they are undone, completed, or broken. However, it will become clear under my framework of analysis that we share many more commitments than Shpall seems to think, and that the elements of

[117] Shpall, "Moral and Rational Commitment," 151.

choice and agent-dependence – which have thus far been central to the concept of commitment – have less to do with which commitments we have and more to do with how we deal with them.

Reconsidering Commitment

As stated early on, I see two general issues with how commitment is understood, both of which affect how the concept may be used in applications like constitutional law.

The first issue is overstatement of the role of agency, or the agent-dependent condition. It's a widespread belief that commitments are more *mine* than other kinds of constraints or bonds, precisely because I have in some way *chosen* them. The commitments I have are thought to be more strongly linked to my agency because they are, in some way, *expressions* of that agency; they're shaped by what I value or what I believe, the goals I've set for myself, or the kind of image I'd like to portray.

This first issue is quite prominent in everyday uses of commitment, where the associations between commitment and positive states like investment or devotion strongly hold, but it is also present in academic uses. While for Elster it makes sense to insist on the agent-centric conditions (given his project is individual agency), it's less clear why such a condition is necessary for Shpall. Certainly, Shpall thinks reference to agency is required to explain how our commitments vary so wildly, but it's not clear why he thinks this wild variation exists in the first place. While there might be great variation in the choices we make, or the things we value, this isn't quite the same as variation in our *commitments*. Instead, this seems to me to beg the question: our commitments only appear to vary widely because Shpall has such a narrow conception of commitment in mind.

The second issue with how commitment is understood is in the *kind* of constraint it is taken to be. Here the problem is conceptually messy (owing, no doubt, to the confusion itself). On the one hand, since the agent-dependent condition requires that commitments are constraints

we choose for ourselves, there's a sense in which commitments aren't *really* limits. This mindset relies on the 'moral magic' of consent[118] to differentiate between constraints that *limit* one's agency and constraints that *expand* one's agency; limits I chooses for myself are of a different pedigree to those which are externally imposed. This explains both the desire to justify constitutional precommitment as a democratic practice – since then there would be no coercion in enforcing such constitutional constraints – and the democratic objections to constitutional precommitment – since constitutional constraints are, in fact, rarely chosen by those under their power.

But on the other hand, there's also an assumption that commitments are *extremely* constraining. This relates to one of the worries that Elster has about constitutional precommitment; part of the reason that constitutional precommitment just can't work is that fact that extra-constitutional action always remains possible; commitments (or, at least, precommitment) must impose some inescapable bind on the agent to be 'true' precommitment. But constitutions and constitutional law simply aren't compatible with constraints of this stringency. They are but 'parchment barriers' incapable of enforcing themselves. If Elster is right that (pre)commitment requires impossibility of reneging or failure, then clearly constitutional "precommitment" could never be. What's implicit in this argument is the thought that commitment is something more stringent – more constraining – than a constitution could ever be.

There are kernels of truth, I think, in each of these issues. An analysis of commitment that exclusively centers on agency would be severely lacking, but so too would an analysis that completely ignored agency. And while I agree with Elster that commitments represent an

[118] I mean nothing negative or dismissive by the phrase 'moral magic.' While I know 'moral magic' carries dismissive or snide connotations, I use it here only to refer to that *je ne sais quoi* - that intuition, however widespread and however justified, that assent is theoretically major and significant.

(almost) inescapable constraint, my understanding of this constraint is far more intangible and abstract than the rope that tied Ulysses to the mast of a ship.

The Constraint of Commitment

Even while Adam's belief in the bible might rationally commit him to believing the world was created in six days, he's not inescapably bound to that belief. He could certainly abandon his belief in the bible and rid himself of any rational commitment to believing its contents. It's not even necessarily the case that he's bound to being rational if he doesn't wish or claim to be. And while Obama might be morally committed to allocating funds to green jobs, he's not *unable* to fail to perform this action even absent the excuse of humanitarian disaster. Even Shpall, who thinks that commitments exert some "pressure on the agent to satisfy them," believes that commitments are nevertheless "escapable." [119]

To the extent that Adam and Obama may fail to follow through, it seems like "commitment is constraint" must be mistaken. But the mistake is thinking that the constraint operates at the level of some individual action or outcome i.e., that Adam *must* believe the world was created in six days or that Obama *must* hand over the pledged funds. As seen above, while even Shpall agrees that Adam and Obama are committed, neither is constrained *only* to those actions.

Instead, I think the constraint operates at the next level of analysis, the level at which an action exists in a certain kind of relationship with other actions – for example, those that necessarily follow, or those that are necessarily incompatible, and so on. Commitment doesn't refer to a given action or state of being itself, but the relationship that action or state has with others. The bond isn't that one has chosen to do or believe something, like Adam believing in the

[119] Shpall, "Moral and Rational Commitment," 153.

Bible, but that the doing or believing of that thing entails a certain set of other actions or beliefs that must or must not follow. The mistake of agent-centric views of commitment is that they look exclusively at how the agent is situated inside of a commitment matrix. But the *agent* isn't constitutive of commitment. The commitment exists between the actions or beliefs, and it's only metaphorically that *people are committed* to certain actions or beliefs (given their entering into the matrix of actions and beliefs). Requirement and commitment, on this understanding, collapse into one.

Commitment is a condition of categories, not of persons. Commitment is a state of relation between concepts – concepts are, quite literally, *put together*.[120] The effects of commitment on persons exists because of how our choices *within* a commitment matrix will alter the conceptual background against which my actions and beliefs operate, as well as constrain the subsequent options open to me. Again, my choices matter when dealing with the commitments I have, but they do not determine the existence of those commitments.

A demonstration: Adam believes that the bible is, we assume, literally true. He is *not* committed to believing that the bible is literally true (though he is, for the sake of demonstration, highly enthusiastic and invested in it). Rather, he is committed to believing that the contents of the Bible are true, which includes the claim that the world was created in six days. This commitment comes about because (it is assumed) Adam is being held to standards of rationality, and these standards contain a law of non-contradiction. It is at present contradictory, and therefore irrational, for Adam to claim that he believes the Bible is true, and yet also reject the

[120] This is consistent with mutually exclusive concepts. Such a relation would be understood, as we know, as a P.-Q or P>-Q form of 'combination.' I'm merely pointing out the etymological parts of 'commitment' as being informative.

claim that the world was created in six days (which is a proposition contained in the Bible). Thus, the form of "Adam's"[121] commitment here is as follows.

1. Adam can reject any claim to being rational, in which case he's not bound by a law of non-contradiction.[122] It's not physically impossible to hold contradictory beliefs. And thus, Adam *could* claim to believe the truth of the bible but reject that the world was created in six days.
2. Adam can give up his belief in the literal truth of the bible,[123] in which case he's no longer rationally committed to believing that one proposition contained therein. He may still have no intention to be rational, but he would not be *irrational* if he abandoned that grounding belief.
3. Adam could decide he *does* wish to be rational and continue to believe in the truth of the bible, and therefore take on board the belief that the world was created in six days. This would perhaps create a new set of rational commitments or place his beliefs in tension with other beliefs he may have. But *this* commitment would be (tentatively) settled.

These are ultimately Adam's only options. If he doesn't care about consistency or rationality, then Adam could ultimately do or believe whatever he'd like, though he may be committed to certain consequences as a result, like having a really hard time navigating a world that claims to operate according to the rules of logic. But assuming he does wish there to be some rational standard by which he operates, his *legitimate* options are as listed.

These are his only options because of how the relevant concepts and categories relate to one another. But this singular fact –Adam's belief in the bible and its relation to the premise "The world was created in six days" – is not an accurate representation of the real state of commitment. When we think about a test case like this, we obviously include the abstract, 'interesting' categories, like 'rationality' and 'truth,' but those aren't the only relevant concepts.

---

[121] I scare-quote this, as it's not Adam's commitment as much as it is the commitment between the rules of logic and the having of certain beliefs. Anyone in Adam's position will have the same commitments. It has quite little to do with him beyond the fact that he is our agent in question.

[122] This, I suspect, is controversial. At minimum, even if one is a 'reasons externalist,' and thinks that Adam is indeed bound to the rules of rationality even if he wishes not to be, I ask if one could concede that there is a difference in the kind of bond Adam has here, compared to whether he did, in fact, also claim to think rationally.

[123] This is also compatible with his belief in the Bible changing. Similar to a point I will make in Chapter 5, Adam may choose to belief the Bible is more metaphorical, rather than a literal version of events.

The *kind* of statement "I believe the Bible is true" commits Adam to believing what is written on the pages *within* the Bible because of how a book relates to its pages, how a page relates to its textual contents, and how textual contents relate to language and truth, literal and metaphorical statements, and so on. The relation between the ideas or conceptual categories – from the nature of truth to the concept of a book – make up and contribute to the matrix of commitment. Adam's case is one illustration of a commitment with a *very* simple matrix of beliefs, yet even this simple example implicates so many categories and concepts that are simply taken for granted.

This has been a very cursory discussion so far, and I'd like to get into some particulars about the nature of commitment. The best way I see to do this at present is to introduce a few distinctions that, I hope, will help flesh out the picture. For clarity and ease of comprehension, I'll begin this mapping out with a clean slate in the next chapter.

Chapter 5: Commitment Without the Agent

The one main thing that I think Elster and Shpall get right is the thought that commitment is a kind of constraint, the normative pressure of which exists independently of one's attitudes. Unlike the folk conception, which tends to conflate the idea of this constraint with things like investment in or enthusiasm about the constraint, Elster and Shpall do well to separate the constraining normative pressure of a commitment from the individual's choices or attitudes. Nevertheless, I think they err in conflating the existence of a commitment from acts of individual agency.

The agent-dependent condition limits commitment in such a way that one's choices become the only source of commitment. In Elster's case, the focus on agency seems plausibly to be a side-effect of his particular research question in *Ulysses and the Sirens* rather than his conceptual belief about commitment (though his statements about precommitment in *Ulysses Unbound* may limit how true this interpretation can be). For Shpall though, as we've seen, agency is front and centre. He insists on a distinction between requirement and commitment, and concludes that commitments are, in a sense, what's normatively 'left over' once all our moral or rational requirements are accounted for. The agent-dependent condition explains why (he thinks) different people's commitments differ so widely; we make different choices, and thus we have different commitments.

But I'm not convinced.

The distinction between requirement and commitment is only needed because of the confusions about agency and commitment, and vice-versa. If commitments are not only those things that we choose for ourselves, then there is no great variability which must be explained. If commitments are not those things we choose for ourselves, it's unclear what we make of the

apparent distinction between commitment and requirements. This distinction, to me, does not seem to contribute anything obviously useful to discussion, but instead seems only to confuse us about what commitment can possibly be (since it must be different from 'requirement'). Shpall wants to maintain a difference between commitment and requirement, but to me, requirements *are* commitments, and thus our commitments don't vary nearly as much as Shpall et al. might think.

Furthermore, it seems like this agent-dependent condition has the relation between commitment and choice backwards (or, at least, artificially limited). Certainly, even on my view there is variation in commitment across people and across time and, certainly, choice has a minor role in explaining some of this variation. But the choices we *get* to make are themselves determined by the commitments we already have, as the commitments that bear on us are what create the conditions of choice. I think I may confidently remind us that there does not exist a rational being who started life from a condition of free choice: we're committed from conception to death. It's never agency all the way down.

Still, something needs to be said about what our commitments are and from whence they come. While I do wish to have a more specific discussion of commitment and agency, I will reserve it for the next chapter for ease of compartmentalisation. This chapter will limit itself to the sources of commitment. I will be introducing further distinctions in later chapters (the "kinds of commitment" and "paths of commitment"). For ease of conceptual organisation, I use the label "two-fold distinction" to refer exclusively to the category wherein I've placed the sources of commitment.

Sources of Commitment: The Two-fold Distinction

What the agent-dependent condition seems to overlook is the fact that choices we make, even where they have constraining power, can only do so when and because they operate against

a particular kind of background. If I'm choosing to quit smoking, then I can choose to leave my cigarettes at home, or to throw them away, to avoid the temptation to smoke. But if my vice is biting my fingernails, I can't leave my fingers at home. I suppose I could amputate my hands, but if my reason for wanting to stop biting my nails is, for instance, to have nicer looking hands, such a measure may, as they say, throw the baby out with the bathwater.

Consider one of the most common reasons given for quitting smoking: "I want to be healthy." We know that the hidden premises here are that smoking is unhealthy, and that the goal of being healthy is therefore incompatible with the continued practice of smoking. Smoking takes a toll on the lungs and arteries and, for someone hoping to improve their fitness or longevity, it's a terrible addiction to have. We can give extensive explanations of why or how smoking harms the body, but we sort of simply acknowledge the fact that smoking is unhealthy and choose whether we want to risk it.[124]

But now consider the fanciful child who says, "then we should make smoking healthy." From a purely abstracted, uninformed position, the logical premise of creating 'healthy smoking' is just as much an option as giving up 'unhealthy smoking.' It's knowledge of the way of the world that *decides for us* whether 'healthy smoking' is a real possibility or not.[125]

It's an almost trivial observation, but it's nevertheless significant. When all conceptual focus is given to the role that agency plays, it's too easy to overlook the fact that agency can only *have* a role because it is working within some pre-existing framework of constraints, relations, concepts, and so on. Agency *is* taking advantage of the order of operations of the world around us, often with the hope of imposing some modicum of our will on it. Shpall would like to deny

[124] Obviously, this is complicated by the addictive properties of nicotine, but for the purposes of illustration, we're assuming agency is intact.
[125] But nor are we *required* to think of smoking as unhealthy. It simply is – it will have the effects it has regardless of whether we try to deny them or doublespeak them away.

that the world or other people can commit us; but if this were the case, there would seem to be little opportunity for or value in agency to begin with.

For me, then, choice and agency are *not* sources of commitment – at least, not in the understanding of *source* as I use it here. By 'sources of commitment,' I mean to pick out the background conditions that constitute or explain *why* certain categories or concepts are arranged as they are. Certainly, wanting to be healthy is a reason why one would quit smoking, but this is a very superficial 'why;' the kinds of 'why' I have in mind refer more to the category of 'why is the pursuit of health incompatible with the practice of smoking.' This kind of 'why' is something much bigger than any individual person's agency; these deal with the way of the world.

Commitment exists – that is, we are constrained in the ways we are – because of the way of the world, but this can be true in two different senses. In the first sense, commitments come about by social practice or construction. In the second sense, commitments come about because of the nature of the physical world. I call these sources of commitments *social* and *natural*, respectively.[126]

Two-fold Distinction I: Social Commitments

Social commitments are relations between concepts and categories that are operative because people treat them as existing, and which would not exist but for this fact. This does not refer to those constraints that are operative independent of human behaviour. The relevant detail here is not *that* we treat these commitments as existing in our social practices, but that they *wouldn't* exist if not for that practice. Among these commitments we might consider those relating to currency and its valuation or devaluation, legal rules and methods of government,

[126] The distinction between natural commitments and social commitments may be understandable as akin to the distinction between natural and social *kinds*, but for lack of full familiarity with that distinction, I do wish to outline these sources independently.

moral concepts and judgments associated with certain actions or values, aesthetic judgments, cultural traditions, and so on.

The establishment of social commitment needn't be an intentionally concerted or centralised effort, though many of the central examples of social commitment *have* come about by deliberate concerted human action: a new criminal law strikes a bond between an action, like consuming a certain plant, and a conceptual category like "criminal" or "punishable" such that henceforth all who consume the plant are 'criminal' or 'punishable' and are liable to the associated consequences and judgements. Our social commitments are frequently generated in this active way. But we may also passively forge these commitments through the slow establishment of tradition: mere coincidence may be granted special social significance, or momentary fads may become standard expectations. There might be an established social rule about which side of the escalator to stand on, or there might come about a decentralised pattern of behaviour as a result of cascading individual actions.[127] Whatever the case may be, the point is that there is only "a right side to stand on" because people conceptually treat such a relationship between these concepts as existing: nothing else binds together the categories of *escalator*, *correctness*, and *standing*.

Given the nature of these bonds being in a sense 'artificial' or socially constructed, it is possible for social commitments to be undone. The bonds between the actions and consequences, etc., are scaffolded only by human activity and thus we (collectively) *could* choose to change their forms or remove them entirely. We could collectively abolish the practice of criminal law,

[127] Example of 'cascading individual action': I stand on one side of the escalator out of fancy, and a person stands behind me on the same side. Subsequent people stand behind *us* (whether thoughtlessly or because they *think* there is a rule about it), and suddenly we have the appearance of rule-based behaviour.

thereby completely dissolving any and all commitments related to criminality and punishment.[128] Commitments related to avoiding criminal punishment, at the very moment that the system is dissolved, are themselves dissolved. Our conceptual matrices would cease to include considerations of avoiding criminal punishment, changing the legitimacy of certain decisions. Conservative attitudes will linger – such as thinking that those formerly criminal actions remain undesirable – and they may remain *morally* unacceptable, but they would importantly cease to be *criminal*. None of this entails that changing or dissolving social commitments is *easy*: political movements, civil and human rights movements, and cultural, legal, and constitutional revolutions highlight how the fact that something *could* be otherwise says nothing about the perilous path to making it so. Rather, the possibility of dissolving these commitments only serves as a way of distinguishing them from natural commitments.

Two-fold Distinction II: Natural Commitments

If social commitments are those that come about from humanity's social and conceptual activity, natural commitments are those that come about *independently* of humanity's social and conceptual activity. Here I have in mind commitments we have because of the literal way of the world, whether that is biology, the planetary cycles, natural forces, etc. Natural commitments are clusters of actions and reactions, or causes and effects, or conjunctions and disjunctions that exist because of the facts outside of human control or agency and that are operative independently of human intention or desire. Among these commitments we might consider those relating to the Earth's diurnal cycle, or the necessary conditions of life, or the facts of physics and chemistry.

Importantly, when I say that these commitments exist independent of human desire or intentions, I do not mean to imply that humanity cannot or does not affect them. Our

[128] It should be clear that if it is a collective commitment, then an individual cannot undo it. I may wish not to be bound by it, but to the extent the commitment exists, I will be forced to reckon with it.

deforestation can wipe out large ecosystems while our conservation can allow species to proliferate. Our medical interventions can slow the development of disease or hasten the development of tissues. Obviously, to the extent that we cannot help but be embedded in these cycles we will both affect and be affected by them, causing certain changes and bringing about certain effects, and thus human agency isn't fully inert. But what we *cannot* do is change the rules of these games: we cannot change that certain chemicals have the effects that they do on the environment or on a tumor.[129] At best, we can exploit the rules for our ends. We cannot prevent the Earth's diurnal cycle, but we *can* use fire and electricity to mitigate the effects of darkness; we can't avoid gravity, but we *can* use tools like parachutes to slow down our descent or powerful jet engines to work against its force; we are limited in how long we can spend underwater because of our need to breathe, but we *can* extend that limit with scuba gear or submarines.

These examples help to delineate what I take to be the difference between social commitment and natural commitment: whereas social commitments, as the product of human creation, can be fully dismantled by human activity, natural commitments cannot. We find novel and highly creative ways of *circumventing* these natural bonds, but the bonds themselves continue to hold.

Both natural and social commitments have a general and a specific application. In the discussion above, I have focused on the general application of each. All (or most) people are socially bound by economic, political, and/or legal structures, and all (or most) people are

[129] What this does not imply is the kind of thought, for example, that so long as humans roll back in our behaviour, the environment will correct or restore itself, etc. This is not the thought that once humans stop overfishing or clear-cutting, certain ecosystems will redevelop as they were. Rather, it's the thought that we don't have control over whether these corrections happen, as it depends on the processes involved. We may be able to artificially support these processes and much of our very technical and refined scientific knowledge allows us to create the ideal conditions for these processes to take place. But we aren't able to change the fact of these processes, and the fact that *those* are the ideal conditions for them.

naturally bound by Earth's physics and cycles. But there are also individualised, non-general, or specific applications for both. Specific social commitments may be those that arise out of promises we make for ourselves, which apply to and constrain just us (and whoever else might be involved).[130] Specific natural commitments may refer to our physical constraints, like disease or disability, that apply to and constrain only us (and others like us).[131] Variation of social commitment exists across time and culture, as things come and go out of 'fashion' or as changing conditions require changes in our behaviour. There are therefore as many different levels or lenses of analysis are there are 'communities' or 'identities' of people.

The Twofold Distinction: Theory Vs Practice

The above distinction is highly abstract. Of course, like so many distinctions in philosophy, a clean dichotomy between natural and social commitment would be too easy, and thus there are caveats and concerns associated with this distinction. I think there *are* commitments that definitively exist in either category, but it seems to me that to think all are either one or the other would be mistaken. It may be better to think of this dichotomy as representing opposing ends of a spectrum, where many commitments fall somewhere along the middle. It seems this must be the case, given how our concepts build on themselves and become highly intersectional.

One theoretical problem targets the distinction itself. To the extent that all conceptual activity is human activity, it seems that all these purported natural commitments are in fact social

[130] I contrast here between the social practice of promising and the making of an individual promise. I discuss this more in the next Chapter.

[131] I think this point holds most strongly in the social case and is maybe not as clear in the natural case. Many people can have the same disease or disability, for instance, and so there's a sense in which it's general (or, at least, not individualised). I need to do more work on this thought, but the general distinction in the natural case might be about comparing what the average person can take for granted- those with 'full mobility' can take for granted that things are "built for them." For someone with limited or assisted mobility, barrier-free access cannot be taken for granted, as it's easy for 'non-standard' mobility constraints to be overlooked by developers. The result is that two people relying on the same transportation system can have very different experiences if, for instance, not every bus or subway station has an elevator or ramp access.

commitments. To use very crude and rudimentary Kantian language, even if there are 'things in themselves' that exist around us, *our* interaction with them is strictly filtered and interpreted through certain faculties of understanding, resulting in some category of 'natural' commitments that are, in fact, social all the way down. As such, the category of 'natural commitment' is empty (or, at best, a misnomer).

I think there's value in this issue as a point of clarification. One difference, generally, may be that natural commitments often play a role as a check on our conceptual activity more so than social commitments. We indeed create complex systems of concepts and categories that we use to understand the world around us, but we also use the world around us to determine the veracity of our systems: when our conceptual systems don't track correctly, we think something is wrong with our system, and not vice versa.[132] I don't mean "natural" to refer to anything ontologically pure or completely free from human interaction; such a thing would be impossible. Instead, I mean to highlight a particular kind of existence that – though understood (and thus interpreted) through and against human conceptual activity, its full 'existence' is not derived from human conceptual activity.

This is different from how we use social commitments, which presuppose a kind of veracity and, on that basis, categorise the world. This is not to say that we don't ever intend for our social categories to track something 'real' in the world, or that we don't ever critique or change our social commitments because of the world around us. The point is only that, where a social commitment or category does not reflect the world around us, it's not (necessarily) because of a mistaken categorisation, but simply a lack of categorical or conceptual belonging.

[132] Except when things are going terribly wrong, such as when efforts at indoctrination or propaganda are at an all-time high.

To demonstrate, consider concepts like those of 'male' and 'female' and 'sex' and 'gender.' There are several commitments associated with these conceptual categories, some natural and some not. It's a matter of certain biological facts that bodies of a particular kind have particular vulnerabilities *owing to their being a body of that kind.* Natural commitments associated with these concepts might be those relating to the differential risks or presentations of illness and disease, to differing levels of certain hormones, or to variable ranges of muscle mass or bone density, while social commitments may relate to expectations surrounding economic or political participation, interpersonal relationship and conflict styles, or methods of communication, etc. The struggle, of course, is that it can be notoriously difficult to know whether some commitment comes about via social construction or nature. "Male," for instance, is generally associated with a lower life expectancy than "female," but it's not clear whether this is something natural, based perhaps on different hormone dominance, or social, relating to cultural expectations about masculinity or political and social participation.[133]

There *do* seem to be biological differences to which the concepts of *male* and *female* are meant to be responsive, and thus the *having* of different categories does seem to be an effort at describing a real set of natural commitments. But there are other differences that, while *thought* biological, are significantly (if not entirely) formed by social construction (albeit, scaffolded on top of natural commitments). Both concepts have some minimal natural grounding that, at some point, were used to underscore or justify gendered social commitments, such as what is expected of a person who is thought to belong to the category, and how those who subvert expectations are dealt with or treated. Given the strength that we know social normativity to have, social

[133] And all this discussion doesn't even get into the question of whether the two exclusive concepts of 'male' and 'female' co-extensively provide a full and complete explanation of the natural phenomenon they're trying to describe.

commitments can persist for extended periods of time masquerading as natural and, therefore, necessary. Thus, even if there are clear differences between natural and social commitments, I do not think it is easy to know, for each commitment in every concept, from whence it came. Nevertheless, it does seem that there *is* a difference between the kinds of commitments that would be called (rightly) natural and (rightly) social.

Distinguishing between natural and social commitment matters, since it can help determine how rigid our constraints are. To be committed is to be limited in certain ways by the operative relations between concepts and categories. While this holds no matter the source of commitment, the source of commitment can indicate how we relate to these commitments. In the case of social commitments, bonds can be fully built or deconstructed based on human activity alone: as the group does, so it shall be, and should the group decide to do otherwise, then those commitments will be otherwise.[134] How strongly tied 'natural' concepts like 'male' and 'female', or 'sex' and 'procreation' are to derivative concepts like "man" or "woman" or "mother", depends on the purposes we want those concepts to play, and the kinds of commitment matrices we are hoping to create. The group can decide to divorce the concept of "female" from certain characteristics or behaviours, which will consequently change who belongs to the conceptual category. [135]

[134] Again, this does not mean that "getting the group to decide" is easy. It only serves to highlight the differences between a "one-and-done" intervention, on the one hand, and the kind of intervention that is required in constant conjunction.

[135] Sally Haslanger's arguments about race and gender, for instance, can be thought of as making the case that the social commitments associated with the concepts of race or gender ought to be more strongly separated from the more natural commitments of skin colour, genetic lineage, sexual reproduction or phenotypical presentation. Haslanger's intention is to create a commitment matrix for gender that centres the particular kinds of gendered subordination that operate alongside (or because of) presumptions about how the social category relates to the natural category of 'female' (or 'sex' generally). See Sally Haslanger "Gender and Race: (What) Are They? (What) Do We Want Them To Be?"

For natural commitments, however, such efforts are inconsequential. It does not matter, for instance, if human beings cease to categorise each other according to chromosomal sex determination as a matter of social practice; there will still be certain kinds of differences between different bodies that bring with them different natural commitments. Those natural facts may cease to relate to any social commitments, like standards of appearance or behaviour, but they may still relate to particular kinds of vulnerabilities, experiences, or capacities. If we erased the concept of health, smoking would yet produce the same effects that we cite as reasons for thinking, under the operative conception of health, that smoking is unhealthy. We can't change how the natural commitments operate, though we certainly can change whether and how they are valued.

There's an important note here. I do not mean to speak of natural commitments as though they *never* change. There are and have been significant changes in the natural world that would seem to immediately call what I've said into question; climate change is one such event, and extinctions may be another. Even beyond our own planet, changes to the natural world are inevitable; there are no natural commitments that are stable in the way I seem to be discussing them, except perhaps at the highest, most universal lens of assessment. I recognise this, and I do not want to be misunderstood as saying "social commitments are ones that change, natural commitments are ones that can never change." It's not about *change*, but about *control*.

Social commitments belong to people; nature does not. The best we can do with natural commitment is to manipulate different natural constraints in our favour. We have certainly done an impressive job at understanding and manipulating nature for our ends, but this does not mean it is something under our control. We do not decide whether there is gravity; we simply learn how it seems to work and decide how we're going to navigate its effects. We do not decide

whether there are seasons; we simply decide whether we're going to travel to avoid them or dress to survive them. But the seasons and gravity *can* change. Winters may become warmer, summers cooler, days longer, as the planet goes through its life cycle. Species may change as generations pass by, eventually resulting in significant variation. Natural commitments aren't static; they're simply uncontrolled.

Sources of Commitment: The Two-Fold Distinction vs the Agent-Dependent Condition

To further explain how I think these sources of commitment work, I want to compare the two-fold distinction model to the agent-dependent condition model and pull out some of the main points of discussion.

One obvious point of difference is that the two-fold distinction does not centralise (or even mention) individual agency. The existence of a commitment bears on an individual agent through their relationship to the natural world or through their relationship to the broader social context in which they live. Whether (and how) one is committed goes beyond the choices one makes, instead depending on the natural and social conditions in which one finds oneself. This complicates the task of determining which commitments we share and which we do not. It's not the case that all human beings share all commitments, as there are natural differences, not only in the kinds of bodies we have, but also in the natural environments in which we live.

Beet and Bot are two people who live in different parts of the world: Beet lives in a temperate climate with fertile land, but a relative lack of fauna; Bot lives in an artic climate beyond the tree line with a surplus of wildlife. We'll assume that each belongs to a small community, but the group is, for my purposes here, insignificant.

Beet and Bot, as two human beings, are going to share many basic biological commitments; they need to consume calories and water for survival, require some kind of shelter

and clothing for protection, and are going to need to sleep occasionally to maintain fitness and sanity. These individual biological facts are commitments in both Beet *and* Bot's commitment matrices. From a biological point of view, their decision-making is the same. But these are not their only commitments.

Beet (and the community) is committed by the fact that hunting is simply not a sufficient source of food for the community; prey in the area is underpopulated, and the risk of hunting to extinction is real. Given the surplus of fertile land, agriculture is the best decision for Beet and the community to make. But agriculture requires a certain kind of rootedness; the community, if agriculture is to be successful, cannot easily pick up and move. Relying on farming will require that the community (as a whole or in part) stay where they are. It will require they are sufficiently close to a reliable source of water, both for consumption and for agriculture. Given the relative lack of control over how much tonnage will be available for harvest – as well as when it will be available – certain kinds of technologies will be needed for food storage and preservation to make the harvest last as long as possible, such as refrigeration or canning. The rootedness of the community makes a large granary or cannery a great option, since the heavy stores can remain in place and be taken as needed, but it may also require some kind of protection from the occasional roving bandit.

Bot (and the community), on the other hand, has a different set of issues. Clearly, the arctic climate is not going to be conducive to sustainable agriculture; even if anything *could* grow in the summer temperatures, it's difficult to rely on farming as a sufficient source of food; the short warm periods and geography of the poor-draining tundra make relying on agriculture as a main food source a risky gamble. Hunting, therefore, appears to be the best choice. But the animals follow migratory patterns. While it would be true that the overall population of fauna

across the entire arctic region at any given time is high, it's easily possible for Bot and the community to hunt their immediate area to extinction if they are not careful. It makes best sense, then, to follow certain animal populations throughout their migratory cycle. This will obviously make settlement difficult or unwise; though few may stay behind, some portion of the community must roam with the animals, otherwise the food source will dry up until the creatures return from migration. If the community is divided, then food would need to be transferable to those who remain in the community, and risks spoiling before the return if the troupe roams too far from the settlement. Thus, if the community tries to settle in place, the range of the hunting troupe may be restricted by how long food lasts in transfer. Certain preservation techniques, like salting and drying, may help extend the shelf life, while others, like canning, would create too significant a burden for the troupe to carry. It might be best then for Bot and the community to be fully migratory; they can rely on transferable or temporary structures during their rest times and follow the herd of animals through their complete migration.

The commitment matrix of a well-nourished individual living in a temperate climate with heavy tree and plant cover are not the same as the commitment matrix of an undernourished individual living in an arctic climate beyond the treeline. But neither is it the case that every person's commitment matrix is fully unique. Were Beet and Bot to be found in the same physical condition, in the same geographical area, with access to the same material resources, then their commitment matrices would align significantly; the surrounding environment would impose upon the two of them similar (if not identical) burdens to bear and similar (if not identical) resources to cope.

But commitment also isn't simply relative. In one sense, Beet and Bot share these domain-limited commitments, even while they are apart; were Beet to move to Bot's community,

the same sets of environmental constraints would apply, and vice versa. What is necessary for Bot living in the tundra will be necessary for all who wish to do the same thing; that is the commitment of that form of life. But in another sense, these different natural commitments are unique to Beet and Bot because they currently occupy the relevant region. Though Beet and Bot are both committed by the form of life one would have in the tundra, Bot's commitment matrix is importantly different from Beet's; both are committed by the harshness of surviving in the tundra, but only Bot lives there.

It may be helpful, then, to understand commitment as logically conditional. To the extent that it is true that living in the tundra will require one to migrate to hunt, then if anyone lives in the tundra, they will have to migrate to hunt. From the need to hunt for food, certain other commitments follow; if one must hunt, then one will have to find the prey to hunt, which itself may require following those prey through their migration, which complicates settlement. This chain of reasoning is true for both Beet and Bot; the only difference is that Bot's conditional has a true antecedent, while Beet's does not.

A commitment matrix, then, is nothing more than a massive network of interrelated conditionals with a variable mixture of true and false antecedents. Which conditions bear on an individual within a matrix depends partly on the nature of that individual; humans have certain commitments *because* we are human, or *because* we are mammals, or *because* we are carbon-based lifeforms. But it also depends on how that nature relates to other natures, such as what it means to be a carbon-based lifeform on a rocky planet, and how that differs from a gas planet.

In some cases, the truth status of the antecedent may change, and this can certainly be because of a choice that the relevant agent has made, such as if Beet or Bot chose to move from their respective settings. Should Beet move to the tundra, then the antecedent in "if Beet lives in

the tundra, then Beet must hunt for food" will change to true; subsequently, the full gamut of commitments related to tundra-living and the hunter form of life will follow. But it's also true that the agent is not the sole factor in whether the antecedent changes; massive changes like natural disaster or global warming could change the conditions of Beet and Bot's respective forms of life.

Furthermore, changes in the antecedent may not be possible at all. All commitments related to being carbon-based lifeforms are applicable to human beings, and we are (as of yet) unable to become lifeforms of a different kind. And it's not just biological fact; social commitments may be incredibly rigid as well. In a state where there are rigid prohibitions on social mobility and different sets of rights for different castes, then certain commitments related to the higher classes can never be applicable to those who do not (and cannot) belong to them.[136] It is true for the lower classes that, were they upper class, they'd have the high-class privileges. But it would never be possible (under this arrangement) for them to change their caste.

Obviously, much that I've said here is in direct tension with Shpall's agent-dependent model of commitment. Shpall thinks an agent's *choices* are the sources of commitment, and that absent such exercises of agency, there can be no commitment. He allows that one might have *requirements* absent any exercise of agency, but he does not think that commitment and requirement are related beyond the fact that both are necessary for an "adequate theory of rationality."[137] Shpall, remember, thinks that commitments come about because of one's *own* activities, and that "another person – or, more generally, the world itself – does not have the

[136] But of course, the full commitment itself can change. Such a constraint comes about from a combination of two facts: 1. No one may change the class into which they are born, and 2. Certain privileges are open only to members of a particular class. Should either fact change, then the full commitment itself changes.

[137] Shpall, "Wide and Narrow Scope," 3. His discussion is limited to rational requirement and rational commitment, but I'm speaking of higher-order categories, which would include moral requirements and moral commitments.

power to commit you to actions, intentions, or beliefs."[138] As such, on his understanding, agency (and its exercise) represent the source of commitment.

But this, to me, seems incomplete at best. Even where we have the freedom to choose from among a set of options, we do not typically get to determine the set of options from which we choose. There is indeed a role for choice to play (and for luck, as well) in explaining variation in two different commitment matrices. I may be free to choose one of X, Y, or Z, but I don't have control over the fact that my choices are X or Y or Z. It seems weirdly shallow to emphasise the importance and power of agency at the immediate level (the level at which I make my choice), and to simply ignore the preconditions that set the parameters of that choice, when it is arguably the pre-existing parameters of my choice that determine the ultimate significance and value of whatever choice I end up making.

Bot isn't *required* to hunt;[139] starving remains an option. Hunting is only needed if one chooses to avoid a slow, painful, uncomfortable death. But if this is the case – if one does hope to avoid a slow, painful death – then food is necessary. If the only reliable (or possible) source of food is hunting (because one lives in the tundra), then hunting will be necessary. As such, if one wants to live, then hunting is *required*; but then requirement has the same conditional form as regular commitment.

This conclusion should not be a surprise; I've been saying throughout this chapter that I view requirement and commitment the same. A concept like "rational requirement" is a bundle of concepts and relations that are put together for ease of communication. We're not looking at some discrete entity in the world whose form is self-contained and waiting to be realised; there is

[138] Shpall, "Moral and Rational Commitment," 154. Emphasis added.
[139] Though, if one is required to live, then one might be required to hunt. I suppose the matter comes down to whether living is a choice, and whether it is 'irrational' to not want to continue living.

no platonic form of 'rational requirement.' Instead, we're creating a category of reference for a thing of some kind that is useful as a tool of communication. Requirement, however understood, is nothing more than the conditions a thing must meet to be classified a certain way; rational requirements are conditions of 'good reasoning', and moral requirements conditions of 'good action'. Different groups disagree on what those conditions ought to be, and those differences will obviously result in different needs for those who wish to be considered fully rational (or to avoid "falling short of some rational ideal"), but the status of 'requirement' stays the same. *Required*, like *valid*, *good*, *ideal, right* and other evaluative concepts, are shorthand words that we use to describe certain bundles of commitments. They are subordinate to commitment, not apart from commitment.

Recall, Shpall's concern about this distinction between commitment and requirements rests on the worry of bootstrapping illegitimate beliefs or (in the case of moral requirements) actions. I quote this section again, as I want to be exceptionally clear in picking this particular issue apart.

> "…it is simply not true that Adam is rationally required to believe that the world was created in six days. This would amount to an illegitimate sort of bootstrapping, whereby one irrational belief sanctions another. [Other] writers suppose, correctly I think, that an irrational belief does not possess such a robust legitimizing power."

Importantly, the notion of a rational requirement does not belong to Shpall; he's drawing on vocabulary from theory of rationality and offering his conception of commitment as a complement to that domain. The concerns of this 'illegitimate bootstrapping' are echoed from others in the theory of rationality. Shpall only reiterates them and hopes to add an additional mode of reasoning in this debate.

For the record, I think Shpall's efforts are more than warranted. Even under Shpall's understanding, commitments are fundamental to rational deliberation. Elster, I presume, would

agree; the whole purpose of precommitment – the whole purpose of taking advantage of the constraint of commitment – is the pursuit of rationality despite our imperfect capacities. I think Shpall was completely right to notice the conceptual proximity of commitment and requirement, but only errs in how he understands this proximity: commitment and requirement are certainly conceptually close, insofar as they belong to the same theoretical domain (as do soundness and cogency), but they do not share any direct relationship (such as soundness and validity). Requirement and commitment for Shpall, it seems, are less like 'two sides of the same coin,' and more like two coins of the same currency.

But Shpall's justification for the requirement/commitment distinction is weak to me, especially when we take into consideration the disagreements in the theory of rationality itself. It's certainly fair to worry about illegitimate beliefs sanctioning further illegitimate beliefs. But that is why we have categorical distinctions between legitimate and illegitimate beliefs (or actions, etc.). There's a strange blindness to the functional role that a concept like 'requirement' plays. An approximate equivalent, for demonstration, would be if Shpall were to insist on separating 'morality' from 'behaviour,' out of concern for some kind of illegitimate bootstrapping of bad behaviours under the heading of morality. We certainly don't want just any behaviour to be thought moral, but that's precisely why we have a more specific category of behaviour which is called moral; this is why we generate theories of classifying actions as moral or immoral. But specifying of a subsection of behaviour as 'good' does not result in those good behaviours being not-behaviour; it simply adds to (the assessment of) the behaviour certain conditions that must be met if it is going to be labelled 'moral.'

In his concern about commitment and illegitimate belief, Shpall seems to equate "rational" with *all* thought and belief, regardless of its rational status. But rationality and 'being

rational' refers to a very exclusive limb of the 'thought and belief' tree. It's not individual beliefs that have legitimating power; it's the rules of rationality, or the rules of 'good thought,' that serve to legitimate or illegitimate certain beliefs or belief structures. These individual beliefs are either legitimate or not depending on whether they pass the tests of 'good thought,' i.e., whether they meet the requirements of rationality. Again, these requirements are not apart from commitment but are an instance of a manufactured shorthand for a given constellation of commitments, a particular wing of the broader commitment matrix.

A concept like 'rational requirement' (or 'moral requirement') is useful. There is, I think, a real phenomenon that Shpall is describing, i.e., the category of legitimate beliefs related to one's commitment matrix. But it seems to me that the conception of "rational requirement" is redundant. If the commitments of rationality impose certain conditions, then one is required to meet those conditions, else one is not rational. There is no need[140] to distinguish between different 'levels' of rationality or morality through the distinction of 'requirement' and 'commitment' in order to talk about good thinking and bad thinking, or good action and bad action. Again, requirement might be conceptually *refined,* but it is not conceptually *distinct*.

The focus of this chapter was to explain the sources of commitment and suggest that agency does not play nearly as much of a 'sourcing' role in commitment as the agent-dependent condition seems to suggest. Certainly, an individual's choices may alter portions of their commitment matrix by changing the status of which commitments presently bear on them (given natural or social facts which they must confront), but agency plays significantly less of a

[140] Not *no* need; it's useful to be able to distinguish between an attempt at rationality and a successful attainment of rationality. But they remain a part of the same branch. Perhaps what I mean to say is that the distinction between levels of rationality do not in turn result in the creation of an entirely different 'rationality' tree.

constitutive role. Choice and agency have much less to do with whether we are committed, and more to do with how we might deal with the commitments we have.

In the next chapter, I want to explain some of the different forms or types of commitment. While the sources of commitment deal with why certain concepts and categories exist in particular relation to one another, forms of commitment refer to these relationships themselves; what the relation between concepts is and how those relationships bear on individual reasoning and action. This will provide additional tools for explaining how evaluative concepts (like requirement) work, as well as outline how individual agency operates within a commitment matrix.

Chapter 6: Agency as a Supporting Character

Throughout the dissertation, I have been strongly objecting to the amount of attention that standard accounts of commitment give to agency when determining *whether* one is committed. Both Shpall and Elster's accounts of commitment centralise agency as the committing factor: Shpall thinks one can only be committed through exercises of agency, while Elster simply wasn't considering commitment and precommitment outside the realm of rational agency. I have suggested instead that commitments come from natural or social sources; certain concepts are created or 'discovered' *ex nihilo* or through observation, which are arranged in a vast network of relationships. These relationships between concepts – whether and how different concepts are put together – are our commitments. We are committed from the moment we are conceived. On this paradigm, whether we are committed has little to do with the choices we make.

But I have also insisted that while I want to decentralise agency as the source of commitment, I do think there is yet a role for agency to play. While the role of agency on commitment is, on the whole, insignificant, agency is nevertheless a valuable concept for explaining how different commitments relate to each other, and how we deal with the commitments we have. It's a mistake to think that commitments only exist where there has been an exercise of agency, but it's equally mistaken to think that agency plays *no* individuating role in our commitment matrix at all. Arguably, it's the fact that we are *not* the source of our commitments that makes agency so theoretically significant.

This chapter, then, will consider how agency relates to commitment. To do so, I introduce several different concepts and tools to make sense of one's commitment matrix, as well as how to describe what an individual decision does *with* and *within* one's commitment matrix.

The first important concept is legitimacy. I posit a conception of legitimacy as it relates to commitment, where it is devoid of inherent moral content. Importantly, I am not arguing for the value of this conception over others; I am reserving the word 'legitimacy' for the task of categorising the different kinds of options that individual agents have when making decisions. I am taking for granted that this is the purpose of the concept of legitimacy in this context.

The next few concepts that I consider belong to what I will call the three-fold distinction, and they deal with the different paths related to one's commitments and commitment matrix. The three paths open to an agent are *completion, breaking,* and *undoing*. Every possible action (or omission) will, in some form, complete, break, or undo one's commitments. I will consider each of these concepts, as well as how they relate to legitimacy.

Toward the end of the chapter, I explain how these concepts come together to describe what goes on when an agent exercises agency, drawing on Elster and Shpall as points of comparison. While commitment is central to making sense of agency, agency is not central to making sense of commitment. It is the particular mechanisms of commitment – the particular forms of commitment, the games of solving broken commitments or successfully undoing commitments – that motivate and make sense of many of our important normative concepts.

Legitimacy

The choices we make when we exercise agency all slot into some form of either completing, breaking, or undoing some commitment in our commitment matrix. From the values we hold, the facts of our environment, and the facts of our physical and biological being, there are only so many options open to us when deciding what to do, and even fewer options that are, in our matrix, legitimate for us to do. For this to make sense, I must first explain the relevant understanding of legitimacy, as this concept is vital to understanding the differences between

completing, breaking, and undoing a commitment, as well as making sense of the precise differences between different forms of commitment.

Importantly, I am not arguing for this conception; I am positing it as the understanding of legitimacy I have in mind as it relates to reasoning about commitment. I do not argue for it as I do not ultimately think that my conception of legitimacy *is* different from existing conceptions of legitimacy as found, for instance, in political philosophy or elsewhere. Instead, my conception of legitimacy can be best understood as underpinning all these conceptions (as well as other 'evaluative' concepts like 'valid,' 'proper,' good', and 'required'), albeit at the highest possible level of analysis.

All concepts, ultimately, straddle some line between two groups of objects: the things which belong to the category picked out by the concept, and things which do not belong to the category picked out by the concept. *Legitimacy*, as a concept, is what we will say of all that is contained within that first category of things: *legitimate* things are things which properly belong to the category picked out by the concept of our investigation, whatever concept we're referring to in our investigation. Legitimacy, at its most abstract level, is understood to mean "according to standard."

To expand on this extremely general definition, it may be useful to break it down according to a number of characteristics. The first characteristic of note is that legitimacy is (1) *evaluative*. By this, I mean to say that legitimacy exists as an indicator of some subject's state of being, relative to some pre-adopted criteria. To say some purported X is *legitimate* is to say that this purported case of X has been evaluated according to the checklist that we've decided is the checklist of the category of Xs. Using the language of the above definition, "This is a legitimate X" is shorthand for "This is an object which belongs to the category of things picked out by the

concept X," where the *concept* refers to the set of criteria against which the object is assessed. If a purported case of X does *not* meet the checklist as outlined by the relevant concept, then we will say it is *illegitimate.*

By categorizing things as being either "legitimate" and "illegitimate" instance of some concept, we can revise and refine both our understanding of the objects of analysis *and* the categories themselves. For instance, if a purported case of X does not meet our criteria of X, one of two things is possible. We may simply conclude this purported instance of X is, in fact, *not* X (e.g. a *purported* bachelor who does not meet the conditions of 'unmarried' and/or 'male' is simply not a bachelor). But we may find that the checklist *itself* is incomplete: while it would seemingly require that we conclude this purported case of X is *not* an X (or, likewise, this purported case of not X *is or ought to be* X), our intuitions consider that to be an incorrect evaluation (such as intuitions about whether a celibate man or a male child is appropriately considered a bachelor).

Second, I understand legitimacy as (2) *derivative.* By this, I mean that the nature of legitimacy is such that the substantive meaning of some purported X being legitimate is derived from the evaluative criteria (i.e., the checklist) – *substance is not contained in the concept of legitimacy itself*. Legitimacy, again, is nothing but "according to standard," whatever that standard may be. "This is legitimate" means next to nothing out of context. It is important to be clear on this point. This does not mean that legitimacy, as a concept, has no *meaning,* but rather that unless one knows the checklist against which some purported X is evaluated – unless one knows what concept relative to which the thing is being assessed – the *meaningfulness* of the declaration "x is legitimate" is absent.

For illustration, consider the practice of awarding medals in competition. Medals aren't meaningless in themselves; we know that if someone has a gold medal, they have earned first place in some kind of competition. However, beyond that, the mere *having* of the gold medal doesn't tell us the full story about *why* the medal was awarded or, perhaps, what the person earned it for. They may have eaten the most hotdogs, run the fastest, scored the most or the fewest points, jumped the highest or the longest, or baked the best pie. The medal has a *meaning,* but it is not in itself completely *meaningful*.[141] One could know everything there is to know about the practice of medal-awarding, and yet still not know the facts about a given instance of medal-awarding.

Legitimacy is like a medal; it does not contain its own full meaning. Instead, it points to the existence of something meaningful to those who are aware of the practice of legitimating things. The mere fact of something's being legitimate doesn't tell us what it is a legitimate instance of. It could be a legitimate authority, a legitimate punishment, a legitimate move in chess, a legitimate excuse, a legitimate concern, a legitimate child, and so on. We can know what it means to say something is legitimate while not knowing the full significance[142] of that evaluation (and thus be unable to appreciate the full implication of such an evaluation).

The final characteristics I will discuss are intimately related. Legitimacy as a concept is (3) *narrowly defined* and is therefore (4) *widely applicable*. Legitimacy is nothing more (and nothing less) than a quality held by purported cases of X when they meet the conditions of the checklist against which all Xs are to be evaluated. Legitimacy is not conceptually limited in

[141] Meaningfulness here is not referring to sentimental value. Moreover, this issue of meaning and meaningfulness is even more salient in the cases of those who are not familiar with the practice of awarding medals for competitions. We as co-practitioners know what a gold medal *means* because we know the game of medal-giving, and even we are left with uncertainly as to the full meaning of a given gold medal. To one who is utterly unfamiliar with the practice any meaning is utterly opaque.

[142] "Significance" in the sense of "what it signifies" rather than simply 'importance' or 'value.'

domain or moral content. A legitimate child is a child who was conceived according to the rules of marriage and family: it has nothing to do with whether they are a *good* child. A legitimate crime is an action which is legitimately considered criminal,[143] not a crime which was committed for a good reason. A legitimate play is a move in a game that is permitted by the rules of the game, not a move that all players are happy with. A legitimate interpretation is an interpretation of, for instance, a work of art that fits with the source material being interpreted. Legitimacy itself contains none of these particular qualities or conditions and may therefore represent any and all of them.

There is one final note on Legitimacy. Legitimacy is *all or nothing* – something either meets all criteria which serves as the basis of the assessment, or it is illegitimate. I do not think there are 'degrees' to legitimacy in any conceptual, theoretical sense. When we speak of something being more or less legitimate than another thing, I understand this to be purely metaphorical. We might describe an argument that is missing one premise as 'more valid' than one that is missing two or three – but this is only because it is *closer* to the absolute of validity, not because validity itself admits of degrees. This is a characteristic I think belongs to Legitimacy proper – things can be closer to being legitimate in several different ways, but if it's not fully legitimate, then it is not legitimate at all.

The characteristics of legitimacy, therefore, are that it is an *evaluative*, *derivative*, *narrowly defined* but *widely applicable, all-or-nothing* concept. "Legitimate" is said of things relative to an independent set of criteria, *not relative to some condition of legitimacy itself*. When we are assessing whether something is legitimate, we will have some categorical conditions of what that thing is supposed to look like, act like, or be like. In the case of a legitimate child, there

[143] Where 'illegitimate crime' might refer to something that, for instance, causes significance offence but isn't criminal, or where accusation of a crime is made but actual criminal behaviour is absent.

would be rules relating to the conditions under which the child is raised or conceived (such as, historically, having been conceived in marriage). Those rules themselves can be legitimate (or, in the case of law, *valid*) too, so long as they are made in accordance with the rules governing the creation of law.[144]

The most important part, and the one which I must insist upon repeatedly, is that there is *nothing inherently moral* contained in that which is labeled 'legitimate' nor inherently immoral in that which is labelled 'illegitimate' *as I use it.* Rather, to say that something is legitimate is to say only that the subject belongs to the relevant conceptual category *and nothing more;* likewise, to say that something is illegitimate is to say only that it does not belong to the relevant conceptual category *and nothing more.* This is not to say that there can be no connection between that which is legitimate and that which is moral, but this relationship is strictly contingent: to be 'legitimate' will only contingently mean "moral" in such cases where the stipulated set of conditions *themselves* include moral content.

The status 'legitimate' does not imbue upon the status-holder anything of moral value. To be legitimate does not *conceptually* equate to being desirable, better, or worthy; likewise, illegitimacy does not conceptually equate to undesirability, lesser, or unworthy. Instead, legitimacy and illegitimacy are labels to refer to whether a relation obtains between, for instance, an end being sought, and a means taken to get there: diet and exercise are both legitimate means of weight loss, as is starvation and laxative abuse; selling belongings and seeking gainful employment are both legitimate means of generating income, as are robbing banks and mugging

[144] Note that this isn't limited to Legal Positivism. Natural Law Theory is in essence making the same claim. The ultimate difference between Natural law theory and Legal Positivism is that the former thinks the criteria of legitimacy are determined by morality and/or nature, while the latter thinks they are determined by social practice.

unsuspecting joggers in the park. If I wish to be free of a nagging neighbour, moving across the city is as legitimate a means to that end as killing them.

Hopefully these examples have raised an eyebrow or two, as this is precisely the reason these examples were chosen. The anticipated response is that actions such as starvation, theft, or murder – as negative, hostile, or undesirable – could hardly be considered legitimate under any reasonable metric, especially when there are 'more legitimate' means of achieving the same ends, like exercise and gainful employment. To put my reader at ease, I agree – but not because these means are *illegitimate* means to the stated ends of weight loss, generating income, or freedom from nagging neighbours. Rather, the absurdity of this assessment is that the decisions above all rest on a single, isolate consideration: I am *only* considering how to lose weight, rather than how to do so in a way compatible with other measures of wellness; I am *only* considering how to get money, rather than how to do so in a way that won't have dire consequences; and I am *only* considering how to be free of my neighbour, rather than how to do so in a reasonable or minimally impactful way.

When we make decisions about how to lose weight or how to deal with our neighbour, we're usually (ideally) considering more than the immediate end sought – this is the role and power of the commitment matrix. If I want to lose weight in a sustainable, comfortable, or controlled way, then starvation isn't going to be a legitimate means of achieving that end, as starving is hardly comfortable, sustainable, or controlled. If I want to deal with my neighbour in a reasonable, legal, and socially acceptable way, then killing them clearly isn't going to cut it.

What often *appears* on the face of it to be a simple means/end decision usually involves a rather complex bundle of norms and considerations – we might say, a matrix - so often taken for granted that they may fail to register as qualifiers *unless and until* they are directly challenged or

neglected. This is especially the case when the decision is framed in terms of what I *should* do or what I *ought* to do; these normative qualifiers are placeholders for a bundle of legitimacy conditions that aren't usually spelled out. Thus, just because you *could* kill your annoying neighbour doesn't mean you *should* – because it's illegal, because it's wrong, because it's an unnecessary escalation, because there are less invasive means, etc.. I may have commitments and constraints that come about because of the kind of person I want to be, consequences I may want to avoid, harms I wish not to cause, or my own thoughts on the appropriateness of some means over others. As such, though in isolation, killing is a legitimate means of being rid of someone, it's the overlapping ends I hope to satisfy which place it outside of *my* realm of legitimate possibility. My commitment matrix bars me from engaging in this behaviour legitimately and would require that I either *undo* certain parts of my matrix, or else I *break* a commitment. But the fact that this is the structure of *my* commitment matrix does not preclude someone else's matrix taking on a different form.

This fact goes a significant way, I think, toward explaining what makes agency valuable. We are not committed *because* we exercise some level of agency; rather we need to exercise agency *because we are committed*. Even while human beings share so much (given common languages, cultures, biology, environments, etc.), we *do* yet have individuality; the variability in individual commitment matrices (as a result of time, space, luck, capacity, and so on) mean that personalised reasoning and judgment is often required if we hope to successfully navigate the commitments which *actually* bear on us as individuals. Others may certainly pass judgment on our decision given what they know (or think they know) about us and our condition, and we may certainly seek advice from others who have some kind of expertise. But it seems only *we*

(could[145]) have the necessary access to the whole commitment matrix from within which we make decisions. This does not necessarily mean that I, myself, am the *best* person to make the decision, nor that I am uniquely able to make the *best* decision. It only means that, unless *I* am aware of, honest about, and forthright with my commitment matrix to others, no one else can make decisions according to the commitment matrix that I actually have. Knowing myself and what I may legitimately do is a task in itself; knowing on behalf of another what they may legitimately do is a whole different kind of challenge.[146]

There are important notes about legitimacy that I want to add here. First, what determines legitimacy is not difficulty or feasibility. Something might be extremely difficult but nevertheless remain a legitimate (though burdensome) means to some end.[147] The occasional person may be successful at, for instance, hiding their intentional starvation, or claim to be performing a food strike, or feign illness to evade any intervention. It would be difficult to physically prevent any person from succeeding in losing weight by starving themselves, given the kind and level of intervention that would be required. It would require near constant supervision and monitoring of food intake and weight, all of which would give rise to incredible burden and invasions of privacy. And even with all these interventions, it would still be unlikely that the interventions would successfully break the bond between starvation and weight loss. To the extent that anyone can yet lose weight via starvation, it remains, at some level, a legitimate

[145] Perhaps we do not have full access to even our own commitment matrix. The point I'm making still stands.
[146] To be bleedingly clear, I am not saying "we are all our own best decision maker." I'm talking about the fact of privileged or externally unknowable information that rests with the agent, and the difficulty for someone *other* than the agent to make a legitimate decision.
[147] The converse is true as well. Something may be generally easy to do and yet be illegitimate if a) it does not hold the instrumental relation to some isolated end sought, or b) it does hold a general instrumental relation to the end sought, but not in conjunction with the other specific constraints. Perhaps this could be thought of as a distinction between general legitimacy and specific legitimacy. So, running is a legitimate means of improving cardio fitness generally, but won't be for certain cases of assisted mobility, such as the use of crutches, canes, or wheelchairs.

option. Thus, in this case, like many other cases, it won't be physical constraints or interventions that successfully change the bonds between means and ends, though they may be efficacious in other cases.

Instead, success will come from intangible constraints, like instilling values that preclude or require certain actions, or by stigmatising different behaviours to attach certain attitudes – desirable or undesirable – to people who engage in those behaviours. If the decision changes from merely "lose weight" to include considerations of health, or strength, or social acceptance or desirability, then it becomes harder (or impossible) for a legitimate means to an end to remain so in light of the compounding sets of conditions. Starvation *is* a legitimate means to lose weight – one who starves will, indeed, begin to weigh less – but it's neither healthy nor sustainable. By shifting the goal from "losing weight" to "losing weight in a healthy and sustainable way" or by combining it with additional constraints like "being honest with those who care about me" I have put starvation outside the realm of legitimacy for me.[148] This is not to say that I cannot starve myself, but only that to do so would be illegitimate unless I abandon my constraints of health or honesty. I cannot make starvation healthy, sustainable, or socially acceptable and thus, I cannot take this particular path toward my goal while claiming to pursue these other ends: one or another *must* go.

This is commitment: it is this type of clustered constraint, and the inability to choose outside of these bonded actions and reactions. My commitment is not what choice I ultimately make, but the necessary set of options from which my choice *must* come. Whether by nature or by a sort of 'social scaffolding,' certain clusters of conditions, actions, and consequences are forced into a relation such that one cannot choose which of the set to confront but must take it as

[148] Again, on the assumption that starvation is neither healthy, sustainable, or socially acceptable, and that those who care about me do not want me to starve myself.

it comes. I have agency to choose (some) ends for myself, or even to choose from among various legitimate means that which is the best fit for me considering the ends I have. But I do *not* get to choose which means *are* means, or which means *are* compatible. What I *may* get to choose is whether I keep, break, or undo certain commitments.

I could choose whether to include conditions of health, sustainability, and social acceptability on my decision to lose weight. But having done so, I am forced to confront set conditions: I can continue to starve myself, and give up these other values (opening myself up to claims of lacking integrity and my own feelings of insufficiency); I can maintain these values, and stop starving myself (forcing me to find some other means or to simply abandon the end altogether); or I can starve myself, but continue to *claim* to abide by these values (thereby requiring deceit on my part, or opening myself up to charges of hypocrisy or dishonesty). I can certainly abandon the goal to lose weight, and thus avoid having to choose of these paths, but I don't get to choose where each path goes. Furthermore, I may not choose a legitimate course of action where there isn't one, and nor am I guaranteed that a legitimate course of action is always available to me.

This is a second important note: legitimacy is not guaranteed. The cluster of constraints and requirements on my behaviour may themselves contradict, resulting in the impossibility for me to choose any legitimate course of action. The more constraints that exist on a particular decision, the more likely it will be that a legitimate course of action simply doesn't exist. My examples above may have generated concern that my quite 'wide' understanding of legitimacy seems to allow anything – including murder – to be considered 'legitimate.' However, true as this may be, it's equally true that my understanding of legitimacy may result in certain decisions having *no* legitimate solution. If the compounding commitments in one's matrix lead one's

reasoning in competing directions, one may simply have to find some way of undoing one or more of one's commitments, or else accept that one or more commitments may have to be broken.

Commitment: The Three-Fold Distinction

Recall Shpall's example about Adam and the Bible. The illustrative purpose of the example is to consider how Adam's rational deliberation relates to the premise "the world was created in six days," given his preexisting belief that the bible is true. The deliberative structure of "Adam's"[149] commitment matrix as it relates to this premise is as follows.[150]

First, as Shpall points out, all that is needed for Adam to avoid any tension in his rational deliberation is for him to accept that the world was created in six days. This premise is, it seems, the natural conclusion of the logical puzzle: I believe X; if X then Y; therefore, I believe Y.

Second, Adam could reject any law of non-contradiction. In this case, it'd be perfectly legitimate (though perhaps odd) for him to claim to believe the truth of the bible while simultaneously rejecting the truth of a proposition contained within the Bible. If he's not ascribing to any recognisable logic, then there's seemingly no reason he can't accept that "I believe X, if X then y, and I don't believe Y." We wouldn't consider Adam rational in this case; or, if he is indeed rational, we may call into question how honest he is being about his beliefs (either about the bible itself, or about the rules of logic).

[149] I scare-quote this, as it's not Adam's commitment as much as it is the commitment between the rules of logic and the having of certain beliefs. Anyone in Adam's position will have the same commitments. It has quite little to do with him beyond the fact that he is our agent in question.

[150] In the interest of full disclosure, this does, it seems, align somewhat with Shpall's views on Adam's options here. One point of disagreement might be that Shpall would perhaps deny that Adam may reject the so called 'requirements' of rationality, while I think this remains an option for Adam. But Shpall would, I think, predominately agree with this arrangement of the options.

Finally, he can give up his belief in the truth of the bible,[151] in which case he's no longer rationally committed to believing the proposition that the world was created in six days. He may still have no intention to be rational, but he would not be *irrational* (for this reason) if he abandoned the grounding belief. It would be perfectly in keeping with "if X then Y" for Adam's final decision to simply be "I don't believe X."

Adam's case illustrates a *very* simple matrix of commitments, and the picture can obviously be far more complicated. Perhaps he comes from a highly religious community who will shun those who do not believe in the bible. Should Adam wish to belong to this community, then he is, in fact, committed to believing in the truth of the bible; this is, after all, a condition of being accepted by the community. Regardless of whether Adam ought to wish to be a part of such a community, to the extent that he has this desire, he's committed to believing in the truth of the bible to do so. If Adam were to abandon his belief in the Bible, then he will be committed to abandoning the community he so desires or lying to them about his beliefs.

Some actions or beliefs, like those in Adam's case, exist in mutually exclusive relations, such that we can only legitimately choose one or the other: Adam can reject his belief in the bible, or he can belong to his community. Other actions and beliefs might exist in a conjunctive relationship, where acquiring one inevitably brings about another: believing in the bible necessarily brings with it the belief that the world was created in six days. These (and more) depend on the source of commitment, the type of commitment and, to a lesser extent, the peculiarities of the agent in question.

Adam's case is one illustration of one kind of commitment with a very simple matrix of beliefs. These sets of actions and states of being include not only those means that bring the end

[151] This is also compatible with his belief in the Bible changing. Similar to a point made in Chapter 5, Adam may choose to believe the Bible is more metaphorical, rather than a literal version of events.

about, but also those means which must be avoided if the end is to be pursued, those that gave rise to the action in the first place, those that must be performed to rid oneself of the end, and so on. These sets of actions and states can be sorted into three paths relative to the end pursued: *completion, breaking*, and *undoing*.

Threefold Distinction I: Completion

Completion refers to having done what satisfies the conditions of a particular commitment matrix (without revision). If it is a task to be done, then completion is the doing of the task; if a promise to be kept, then it's following through on that promise; if an action not to be done, then it's abstaining from that action. Completion would be for Adam to accept the belief that the world was created in six days, given his pre-existing belief in the bible. Changing his belief in the bible would not complete the particular commitment, but rather would undo it (which I will explain later). The relationship between the rules of logic (specifically the law of non-contradiction), the contents of the Bible, and Adam's belief in the truth of the bible, is such that the only *legitimate* way of completing (or satisfying) this commitment is for Adam to accept the belief that the world was created in six days.

Completion is not conceptually identical to legitimacy. Not all actions that satisfy commitments are legitimate. Legitimacy is a holistic assessment that takes into consideration the *full* decision matrix and results in judgment about which actions or outcomes, if any, may satisfy all relevant or implicated commitments. Completion, on the other hand, is not necessarily holistic. Thus, we can complete commitments illegitimately, for instance, by acting contrary to the standard procedure or by acting against what we claim to stand for. Completing some commitments may come at the expense of some others, which results in an illegitimate outcome.

In a very simple, very singular case, legitimacy and completion may be identical: the very simplified Adam example is one such case. But where there are many different commitments and

conditions on one's decision, it's not the case that *all* possible means of completion are legitimate (again, relative to the entirety of the commitment matrix). For example, I only have $20 until Sunday but need repay two $20 debts by Friday. There is (in this case) no *legitimate* way for me to pay the full $40 of the combined debts. I may choose to pay the one, or the other, but not both. Should I pay the one debt, then I have completed that commitment; the other debt remains unpaid. I may have no choice but to *break* one or the other commitment.

Threefold Distinction II: Breaking

If I have broken a commitment, then I have behaved in such a way that I have not completed a commitment *which has not been undone*; the commitment *remains* and remains *unsatisfied*. It's going against what one has promised to do or disobeying a direct order. "Breaking" is also embodied in the case of Adam failing to adopt the belief that the world was created in six days despite endorsing the literal truth of the bible and claiming to be rational. The relation between these three constraints – the commitment – is such that they cannot be co-completed: either Adam accepts the belief that the world was created in six days, rejects the literal truth of the bible, or he is not rational (and thus, may not truthfully claim to be so).

Breaking, like completion, can happen in two ways; either I can act in direct contravention of the undertaking by doing what I've been explicitly told not to do, or I can fail to act when action is expected or required. Whether one breaks a commitment though activity or passivity depends on the nature of the commitment in question, and thus can be highly unclear. In some cases, the fact that I've broken my commitment may be obvious: when I promise not to eat your last slice of pie, and I eat the slice of pie, I have clearly actively broken that promise. Likewise, if I resolve to do all my laundry *before* I go on vacation, and I have one remaining load by the time I leave for the airport, then I've failed in that endeavour.

Part of what makes these cases 'obvious' is that the constraint and scope have been clearly defined: in the first case, it's clearly stated what behaviour is expected and in the latter case, I have a definitive deadline. As such, assessment of whether I've broken my commitment is straightforward; I ate the pie despite promising not to do exactly that, and I didn't finish the laundry before my vacation. However, knowing whether a commitment has been broken (or, perhaps, just hasn't yet been completed) isn't always obvious. Many end-points are loosely or generally understood, and their contents may have variable interpretations – one may understand "doing the laundry" to include "putting clean laundry away," while another may consider that a separate task. For the former, not putting laundry away is failure to finish doing the laundry; for the latter, the laundry is done.

Outside of these concrete examples, things get even messier. How do we know whether we've served the value of equality or equity, or whether justice has been served? How do we know whether a particular outcome is fair? Just as difficult as it can be to know when one has accomplished what one set out to do, it can be difficult to know when or whether one has failed. The complexity here isn't necessarily in 'breaking a commitment,' but rather in the (poorly) communicated expectations that make up the commitment itself. This is, as we'll see, a common problem.

Of the three-fold distinction's three components, breaking may be the most difficult to understand. This is not because it is itself particularly complex, but because *breaking a commitment* – or even the word 'breaking' – is usually so normatively charged in ordinary language. Typically, we think of this kind of thing as akin to someone breaking a promise, or otherwise disappointing some expectation, which usually presupposes some kind of moral failing. However, like illegitimacy, I do not think breaking commitments is inherently wrong and

want to emphasise the space between *breaking a commitment* and *engaging in immorality*. I have broken my commitment when I fail to complete the laundry before I go on vacation, but I haven't acted *immorally* by doing so. My resolution has been disappointed, and I may be annoyed by the fact that I'll have even more laundry to do when I return home (which my rational deliberation will now need to account for). But my failure here is a far cry from moral failing.

Breaking a commitment is always illegitimate, but this does not mean breaking is always bad or immoral. Legitimacy, again, is simply a holistic assessment of what will complete or satisfy the matrix of commitments. When not all commitments in the relevant matrix are satisfied (in other words, when some commitments are broken) the outcome is illegitimate. It is possible, depending on one's commitment matrix, that there *are no legitimate courses of action* – a true moral dilemma is one such instance. Different decisions may complete some ends or others, and there may be no option which will complete all of them. One must either undo one or more of the commitments (if possible) thereby eliminating the conflict, or simply accept that one cannot but act illegitimately – some commitments will be broken in the pursuit of the others. But importantly the inherent illegitimacy associated with breaking a commitment does not conceptually impart any inherent moral value or disvalue; whether something *illegitimate* results in something *immoral* depends only on whether it has contravened a moral commitment in one's matrix.[152]

[152] This allows, I think, for internalist and externalist interpretations of moral reasons. I might think my action was legitimate, since I lack the specific interpretation of some moral value. But you, as one who holds that particular interpretation of the moral value, may see my action and reasoning and conclude it to have been illegitimate.

Threefold Distinction III: Undoing

In an earlier chapter, I noted that Shpall's claims of normative difference between moral and rational commitments rests on a tilted comparison between one commitment which has been broken and one which has been *undone*. Undoing refers to a dismantling of the expectations associated with the undertaking in such a way that the goal has been neither completed nor broken. It is, I think, mostly used in reference to the commitment matrix of a single person, but it is also possible for the more general commitment matrix to become undone as well – at least, for social commitment.

To demonstrate: Adam is only ever committed to believing that the world was created in six days because he believes that everything the bible says is true, and (we assume for the sake of the example) because he wishes to be considered rational. But if he were to change his mind about the bible or being rational, then he's no longer committed to that belief,[153] as the original conditions that gave rise to the commitment in the first place cease to exist. Suppose every version of the bible Adam has ever read has not contained the Book of Genesis or has had a misprint. Or suppose, on the basis of some historical research, Adam believes the relevant section of Genesis was mistranslated. Were any of these conditions true, then Adam's commitment matrix would change for him in such a way that the previous conditions of legitimacy also change.

The above example, I would say, is a description of undoing that occurs in the specific commitment matrix for Adam. But the matrix could change more generally. Suppose logic (as 'the' system of rules for rational thought) abandoned a hard rule of non-contradiction; Adam himself may have done nothing to alter his personal commitment matrix, and yet move from a

[153] Though he may be committed to the consequences of irrationality.

condition of illegitimacy to legitimacy because the rules of logic themselves have changed from underneath him. Since contradiction is no longer irrational, Adam may believe in the literal truth of the bible, and stubbornly reject its proposition that the world was created in six days, *and* claim to be rational; now, though, he may do so *legitimately*. Changes to our conceptual matrices can occur completely independent of our doing (and potentially without our knowledge).

Undoing a commitment usually requires that the conditions under which it was generated no longer obtain. This applies in the case of having to take back a promise, such as Shpall's Obama example. The conditions under which Obama promised to allocate funds were, we will presume, conditions of relative calm and stability. After a large-scale humanitarian disaster however, priorities and needs change, such that it may no longer make sense to think the background conditions of such a promise continue to obtain. The reason Shpall is willing to conclude that Obama has not acted immorally by not fulfilling his promise is precisely because the conditions under which the original promise was made have changed. We are willing to 'forgive' Obama for his 'moral wrong' in this emergency because, it seems, a moral wrong was never actually committed.[154]

Like breaking, undoing is somewhat simple in form but can be highly complex in practice. Knowing whether a commitment has been successfully undone or whether it has been broken is hardly easy. There may be disagreement on what the initial background conditions were, whether the conditions have changed, and what is required for a commitment to be undone. If I owe someone $20 by Friday, then I'll clearly break that commitment if I fail to pay them

[154] Alternate take: we might consider a moral wrong to have occurred if we think that certain kinds of promises ought never be made, such as the kind of promise that one has little control over. I could see someone who condemns Obama, not for *reneging* on the pledged funds, but for pledging them in the first place (or on such an advanced timeline, or at such a large amount of funds). Such a person may think that Obama's initial pledge ought to have taken into consideration the possibility of a humanitarian disaster, and his failure to do so means he is bad at playing the promising game. But the point I'm making here remains; there is no issue of a *broken* promise here.

back $20. But what if I pay them back Saturday or Sunday? This, again, will be variable: some might conclude that, given the precise deadline of Friday, I do indeed break my commitment by not paying until Saturday morning; others might be willing, given low stakes or knowledge of circumstances, to recognise the later payment day as completing the commitment. In some circumstances, like law, conditions of completion, breaking, and undoing may be clearly outlined and deliberately unrelenting; in other cases, conditions may be more flexible (and therefore, difficult to quantify or describe).

It may be important to highlight that undoing a commitment is not inherently more moral or acceptable than breaking commitments. Again, these concepts do not have that kind of normative weight. Any relative 'virtue' of undoing commitments rather than breaking them – if there must be one – is out of some perceived value in consistency or integrity only. It is possible that one's commitment matrix may commit one against breaking certain commitments (if, for example, they wish to avoid the associated consequences), and thus, there may be individual or perhaps socially proscribed value in not breaking commitments. If, perhaps, one strives to embody a certain conception of integrity, then breaking vs undoing commitments may have a certain moral quality relative to that end. But I do not attribute this moral quality to breaking or undoing in themselves. Rigidity of commitment may be just as troublesome as fluidity.

Threefold Distinction and Sources of Commitment

I wish to speak on the relationship between the sources of commitment (twofold distinction) and the threefold distinction and legitimacy.

To review, commitment is a rational construct, insofar as it has to do with how concepts – as our tools or building blocks of thought – relate to each other. Again, commitment refers to the relations between concepts, *not* to some level of emotional investment or perception of value. Things are committed when they exist in conceptual relation to each other.

These conceptual relations can arise either because the concepts themselves were constructed to be held in a certain relation, or because they were created to describe things that appear to independently exist in a conceptual relationship: the former category are social commitments, and the latter are natural commitments.

Social commitments refer to the relationships between concepts that are arranged (in a sense) artificially: if humans did not put those concepts in their respective relationships, then the categories that they describe would, in a sense, not exist (nor would the relations between them). This is not to say that the relevant behaviours or actions would not exist, but that the judgments or conceptual understandings attached to those behaviours and actions do not arise 'in nature'. Nature, for instance, does not make an action *criminal,* even if it makes the action. The concept of *criminality* is a social construction created by humans.

Natural commitments, on the other hand refer to conceptual relationships where the conceptual relationship is prior to the conceptual descriptions. There is something we want to describe or understand, and our method of doing so is through conceptual creation. Just as we create concepts in the case of social commitment, we create concepts in natural commitment. The difference is that, in the case of natural commitments, our concepts can be incorrect.

The threefold distinction, recall, makes sense of how we deal with the commitments we have: given the conceptual relationships that bear on our decision-making, certain courses of action will either complete, break, or undo some of our conceptual bonds. If a certain course of action will complete all commitments in my commitment matrix, then that action would be legitimate. If there is some commitment in my commitment matrix that would be broken, then the course of action would be illegitimate (though not necessarily immoral or wrong). In some

cases, illegitimacy may be avoided if the commitment can be undone: insofar as that conceptual relation ceases to exist, there exists no commitment to be broken (or completed).

The threefold distinction, then, exists independently from the twofold distinction. The latter refers only to where the conceptual relationships come from, while the former refers to the individual (or specific) cases of reasoning relative to the commitment matrix that exists. But while the two sets of distinctions are independent, they can interact in meaningful ways.

First, there's the interesting possibility for social commitments themselves to be fully and completely undone. I discussed this briefly before: insofar as social commitments exist as a result of human conceptual activity alone, they can be fully undone on a collective level. Here, again, the individual and the collective come apart.

Let's use law as the example for social commitments. If we are looking at criminal law from the individual point of view, then a single individual is unable to undo the commitments of criminal law. As Austin once wrote

> Suppose an act innocuous, or positively beneficial, be prohibited by the sovereign under the penalty of death; if I commit this act, I shall be tried and condemned, and if I object to the sentence, that it is contrary to the law of God . . . the court of justice will demonstrate the inconclusiveness of my reasoning-by hanging me up, in pursuance of the law of which I have impugned the validity.[155]

I cannot choose for myself to undo the commitments that criminal law creates – but the *collective* can. Should the larger relevant group decide to change that law, or even to abolish the criminal law altogether, then the relevant commitment will have been undone: that innocuous action in question will no longer be conceptually bound to the concept of 'criminal' or 'condemnable', and reasoning about that action will change in every commitment matrix:

---

[155] Austin, *The Province of Jurisprudence Determined*, i85 (Library of Ideas ed. 1954), cited in Hart, "Separation of Laws and Morals." 616.

"engaging in that innocuous action without breaking the law" will change from *illegitimate* to *legitimate* for everyone's commitment matrix.

This is not the case for natural commitment; it doesn't matter if all of humanity ceases to buy into certain natural commitments: they will continue to obtain. But while the existence of natural commitments is, as previously said, outside of all human control, there is still a meaningful interaction with the threefold distinction for our reasoning.

We cannot break or undo natural commitments, since those commitments exist outside of our control. What appears, in our reasoning, to be a broken natural commitment is, in most cases, a set of descriptively inaccurate conceptual tools: the phenomenon we're trying to describe does not align with the tools we're using to describe it. It's not the case, for instance, that the nature of disease itself changed when our paradigm shifted from the miasma model to germ theory: the phenomena were always operating the way they were operating. Instead, what changed was the reliability and accuracy of our conceptual tools: germ theory did not have the apparent failure of the miasma model and was thus deemed a better framework.

This interaction between natural commitment and the three-fold distinction, then, is extremely useful for our natural sciences. The fact that the natural commitments themselves are not within our control offers our conceptual work some constraint, since it provides some independent tests of accuracy for our conceptual tools: even if the tools we're using to make sense of the natural world are themselves entirely human creations, it's still the case that the thing they're meant to describe provides limits on what those concepts can and cannot be. And if it seems like we're breaking some natural commitment, then we've done something wrong. To fix this, we change our concepts to better describe the natural commitment; we have not undone the natural commitment itself.

Threefold Distinction: Theory vs Practice

Thus far, I have used relatively easy examples of completion, breaking, and undoing. But I hardly mean to suggest that it is always so easy in practice to recognise or categorise actions or states of being as those that will complete, break, or undo.

One complication is knowing where to place the margins between completion and breaking in long-term or cumulative goals, like quitting smoking or getting more active. If I, as a smoker, decided today that I am going to quit smoking, how do we know if I have completed this goal? On one interpretation, we couldn't know until the moment I died, as only then would we know that I, in fact, will never smoke again. But this seems an odd way of understanding it. This would require that even after 20 or so years of 'success', I could still only really claim to be "quitting." On the other hand, someone who claims to have quit smoking upon finishing their most recent cigarette hardly seems to understand what it is to quit smoking. These long-term or cumulative cases certainly *have* success conditions, but what those success conditions are is not clear.[156]

A related issue (again mostly with long-term or cumulative ends) is what to do with conditions of 'relapse' or recidivism. Suppose I have success in quitting smoking for several years, but one day, for whatever reason, I have a cigarette. Have I failed to quit smoking? It seems that I obviously have, given that I've done what is the explicit opposite of "not smoking." At the same time, surely at least some failure is an expected part of the journey. We know quitting smoking is difficult and I should think this is because we *expect* there will be slip-ups. But how many 'slip-ups' are consistent with completion? A tentative (and perhaps dodgy)

[156] Perhaps the success conditions of quitting smoking are less about how much time has passed but are more about the quitter's mindset in relation to smoking. Someone who has 'only' stopped smoking for a month, but who doesn't think about smoking even when in the company of others who are smoking might be deeper into their journey than someone who continues to struggle with temptation after decades of success.

answer is that this might have to be decided on a case-by-case or practice-by-practice basis. Some ends may be understood to be inconsistent with any recidivism, while others are looser in their constraint or severity. Discussions about success (or failure) conditions may clarify one's exact aims, which may be revealed to be partly determined by the other aims on one's commitment matrix. So, for example, my resolution of quitting smoking may be informed by my other resolution to be healthier, or by my resolution to save money. To the extent a single slip-up isn't likely to be very harmful to my health 'in the grand scheme of things,' that commitment matrix may be more accommodating than the saving money matrix, assuming cigarettes always cost money.

A final point on this distinction, which will lead nicely to the next set of distinctions, harkens back to something I said about undoing commitments. I had noted that it is possible for Adam to maintain all his same ends and beliefs, and yet move from a condition of illegitimacy to one of legitimacy if the rules of logic changed. Our commitment matrices exist independently of us and what we, as individuals, do. This is quite different from both Shpall and Elster, for whom one's commitments are determined by oneself. Shpall explicitly denies that we can be committed by 'the world,' but I am claiming that we are committed by the world almost exclusively. Our choices may change the precise linear makeup of commitment matrix – since what I choose now may add or remove options down the line – but the amount of control one extends over this is limited. I may choose to be rational, but I do not get to choose what that implies for me, now or in the future. I may choose to break the law, but I do not get to choose whether I'm punished for it now or in the future. I may choose to hold certain beliefs, but I do not get to choose how others react to that now or in the future. Nevertheless, to the extent that I have those beliefs, break those

laws, or wish to embody rationality, I'm committed by the world as it is, not as I think it ought to be.

This chapter introduced the elements of agency as I see them working in this framework of commitment. While we (as individuals) lack agency over the sources of our commitments, we do exercise some agency when navigating the different courses of action that are open to us relative to our particular set of commitments (or, our commitment matrix). In the next chapter, I wish to highlight how this series of conceptual parts I've been creating can be used to describe a series of (what I call) forms of commitment – recognisable and distinct patterns of commitment reasoning. While the sources of commitment make sense of where the conceptual bonds come from, and the threefold distinction and legitimacy make sense of the different ways an agent can engage with those conceptual bonds, the forms of commitment demonstrate how the nuanced relationships between concepts can radically change one's set of legitimate actions.

Chapter 7: Forms of Commitment

Thus far, I have introduced a large number of discrete conceptual parts, all in service of making sense of what commitment really looks like – both in its origin and in its form.

The two-fold distinction – natural commitment and social commitment – indicates the source of commitment. Unlike the agent-dependent model, which nominates the exercise of agency as the source of one's commitment, the two-fold distinction instead attributes the existence of commitment to a social or natural source: certain categories and concepts are committed either because of the laws of the universe, or because of social construction. The commitments we have are operative because they were determined by natural or social rules, and depending on which source they have, will be more or less under our control to change or undo. As our concepts become more complex, so too does the task of differentiating which commitments are social and which are natural. Most, I'd say, are composites of both.

The threefold distinction – completing, breaking, and undoing – is indicative of the state of being committed, where one cannot choose to exist outside of certain conjunctive or disjunctive ways of being or doing. It is the nature of these kinds of bonded actions – those that complete, break, or undo the commitment – that one cannot but end up doing one of them. Even omission results in some outcome, and that outcome will belong to one of the categories. So long as a person lives, they will pass through the series of events, causes, and effects. The three-fold distinction is useful for making sense of individual agency; though we cannot choose what our commitments are, one can often choose whether or how to complete, break, or (in some cases) undo them.

Finally, the concept of legitimacy (as it relates to this commitment framework) is needed to make sense of the practice of sorting actions or decisions into rationally or pragmatically

distinct categories. The parts of the two-fold and three-fold distinctions are necessary for the descriptive accounts of our commitments; part of our reasoning relies on an accurate (or accurate enough) making-sense-of the commitments that actually bear on us. But legitimacy, as a holistic, evaluative, and derivative concept, is necessary for making sense of reasoning about one's commitment matrix as a whole: such a concept is needed to make sense of how we sort different possible judgments into 'is' and 'is not,' or 'ought' and 'ought not,' as well as how we know when and whether to revise our commitment matrix to amend these judgments.

In this chapter, I want to provide some examples of what I call *forms of commitment.* A form (or type) of commitment is a formal or structural commitment pattern that we can recognise, follow, and reinforce in many of our own goals or collaborations. They are, in a sense, commitment games; the forms have certain rules and conditions that apply to all commitments of that type. Different forms of commitment that share the same language may nevertheless result in different conditions of legitimacy owing to the particular interactions of the formal rules and that which each different form takes as its ground of commitment.

I suspect there are many different forms of commitment, given the number of complex social games that humans can play – with each other and with ourselves – but I limit myself to three: *resolution*, *fidelity*, and *deference*. I think focusing on these three will be useful to show both how the conceptual tools I've put forward combine to build highly complex games and can also be used to explain our legal and political practices which seems to rely quite heavily on all three. There are many core disagreements in political and legal philosophy which can be interpreted to rest on disagreement about the forms of commitment at play in a given issue. As such, this discussion may be illustratively useful in the next chapter, where we will return to discussions of constitutional precommitment and democratic legitimacy.

Resolution

While I can make no claim to have fully exhausted this line of thinking, I feel relatively confident in claiming that all forms of commitment presuppose some grounding *resolution,* insofar as resolution, as a means-end form of thought, is the kind of thinking and planning that commitment requires. As such, commitments of resolution seem to be the obvious first subject of discussion.

Resolutions, as a form of commitment, take as their grounds[157] a particular *end* to be achieved or avoided. Just like the New Year's tradition, when we make a resolution we decide to aim for some goal or end, such as quitting smoking or being a better runner, which serves *as* the resolution. But unlike the New Year's tradition, commitments of resolution are not limited to effortful, life-changing decisions. All orientation centered on the pursuit of a particular end – from a recipe to a relationship – is resolution.

Importantly, when thinking about resolution and commitment, *the resolution itself* – the 'not smoking,' the 'being a better runner' – is not the commitment, though it serves as the grounds of the commitment – the reason *why* one has the subsequent commitments one does. The end chosen will be associated with a necessarily limited set of possible means of completion; one can only accomplish a particular end by doing that which gets one to the end. The resolution to quit smoking can only be achieved by an eventual stoppage to smoking; and being a better runner can only be achieved by, first, determining what the metric of progress will be, and demonstrating improvement according to those metrics. In other words, the end sought merely sets the parameters by which that end may be achieved. "Consistently practicing" and "increasing cardiovascular stress" are the commitments; "Being a better runner" is the resolution.

[157] Here, I use "grounds of commitment" in (I believe) the same way as Shpall.

There are some reasons why my general discussion of commitment thus far has predominately been through examples of resolutions (or, discussing commitment in relation to certain ends). First, what I call resolution seems to closely resemble Elster's understanding of precommitment and thus it made most sense to continue the discussion in that vein. The kind of reasoning that Elster's imperfect rationality seems to be trying to explain resembles the same kind of means-end reasoning which is, I think, well captured by resolution. But it also seems to me that resolution simply is the most common form of commitment, both because other types of commitment presuppose some other, primary resolution, and simply because individuals make many, *many* resolutions.

Though the easiest and maybe most accessible examples of resolution are, as I say above, New Year's resolutions or life-changing goals, I tend to think of resolutions as extremely mundane, ranging from cleaning the house to making pasta for dinner. Many of the resolutions we make don't register *as* resolutions, let alone as grounds of commitment, partly because of the tendency to conflate commitment with decisions or obligations of particular significance. If one thinks commitments are only those things that we think are valuable and worth time and effort to bring to fruition, then the things one 'properly' thinks to be commitments are going to have to be things of greater value: our careers, our families, and our moral and political perspectives.

The mistake is that the *end* – being a better runner, quitting smoking, parenting– is *itself* taken to be the commitment, rather than the ground of the commitment. In the case of resolution, the grounds of commitment – the thing that *gives rise* to the commitment – is the end to be pursued or avoided. In the case of being a better runner, *running better* is the end, but what commits me is *what it will take to get there* – improving my cardiovascular health for better performance and stamina, having the time, money, and energy to practice, having the mobility

and musculature to run, knowing what 'good running' looks like, and so on. In even as simple a resolution as going to a café, there are numerous conditions which must obtain for that end to come to fruition, and I *cannot* achieve the end *unless and until* I follow through on those conditions.

As previously discussed, all decisions about commitment serve either as a means of completion, breaking, or undoing. For a given resolution, the number of options available in each category varies: one resolution may have many means of completion and few ways of breaking or undoing; another may have only one means of completion, many ways of breaking, and few ways of undoing. This is, of course, compounded when we take into consideration an individual agent's commitment matrix, which will add in the determinant of whether certain options are legitimate or not: depending on one's particular health constraints, for instance, certain kinds of 'good running' may be illegitimate for me to pursue. This variability presents opportunities for agency, even if the particular resolution I'm attempting to complete is not one which I've chosen for myself, or if I have a number of unchosen constraints that otherwise bear on my reasoning. I may have little choice but to seek gainful employment, but there may be many different versions of what completion of this goal looks like. My conception of what 'being a better runner' looks like may never feasibly be 'win a marathon,' but I may nevertheless find some conception for me, given my constraints.

Fidelity

With one type of commitment on the table, I can contrast it with a second: commitments of fidelity. "Fidelity" often comes up in discussions of marriage or relationships. In closed, monogamous relationships, it is an expectation that partners do not engage in extramarital affairs; in the event of such an affair, one is said to have committed infidelity. However, fidelity is a quality outside marriage: when an artist has produced a stunning likeness, they are credited

with creating a 'high-fidelity' portrait; a translator, accurately capturing the original sense or meaning of a text, is credited with producing a high-fidelity translation; audiophiles, hoping to hear their music as the producer intended, will opt for 'hi-fi' speakers and headphones.

From these ordinary uses, we can glean a common thread: 'fidelity' picks out some source material, whether it be a melody, a visage, or a vow, which serves as the grounds of commitment. Something is said to be 'high-fidelity' for its accurate reproduction of the thing to be reproduced. Doing something in a fidelious manner is not simply doing that thing: built into this form of commitment itself, there are certain prescriptive values that determine the legitimacy of the course of action one takes. I don't just translate a text; I must translate it *accurately.*[158]

Unlike commitments of resolution, which can permit variable means to achieve a given end, commitments of fidelity are more stringent in what is legitimate for a given task. If my task is to provide a translation of a text, I am quite bound by the source material. It does not matter if, amid my translating, I find a more concise way of explaining the ideas or a better order of explanation, nor does it matter if I come to find that the subject matter I'm translating is inaccurate or false; what matters is that I'm accurately reproducing the original text, in all its inefficiencies and errors.[159] Fidelity tracks the difference between how we assess someone who claims to be *translating* a text versus someone who claims to be *explaining* a text: the latter (though still bound by a certain level of fidelity, for sure) has a significantly wider parameter for completion; *taking inspiration* from a text has an even wider parameter. In the physical arts,

[158] Some might consider this to be a redundancy contained in the concept of 'translating,' such that to translate something in any way but 'well' isn't to translate at all. I say this seems to ignore the fact that we very much do discuss better and worse translations, especially in philosophy, and that even in the category of 'translated' (meaning accurately capturing the original meaning of the text) there can be greater and lesser fidelity to what the original author is thought to have meant given what they say.

[159] This might vary, depending on what the particular task of translating is. But the point remains – if there is some literal or spiritual meaning which is to be maintained, my options are severely limited by what it presented to me to begin with.

restoring a bust of Aristotle, *creating* a bust of Aristotle, and creating a bust *inspired* by Aristotle – these three ends have widening parameters, such that what may be legitimate for the second and third end may not be for the first. The thought is that fidelity requires a more accurate recreation of the source material, such that the accuracy of any attempt is subject to evaluation against the source material itself.

All commitments of fidelity presuppose a particular resolution, since there is some end I hope to achieve. My resolution may be, for instance, to paint a portrait of Beet. But given the nature of this resolution – of the end I seek – there is a presumed kind of constraint that is not present if I were simply resolving to paint a picture or if I were resolving to paint any old person. Given the particular target of my resolution – Beet's appearance – there is a presumed constraint on my pursuit of the end and whether or not I could be said to have achieved it. On top of the commitments I have in virtue of what is required to *paint*, I am further committed by what Beet actually looks like; these further commitments relating to Beet's actual image are the commitments of fidelity. If my painting does not look like Beet, then I will not be said to have completed my resolution, no matter how good my painting may otherwise be. There is only one way I may ultimately accomplish the goal: my painting *must* look like Beet. If my painting does not look like Beet, then I have failed to complete my resolution.

In everyday life, I think the prime example of commitments of fidelity are promises or contracts. Like Shpall, I think promises *do* have a central place in discussions of commitment though, unlike Shpall, I do not think that they are wholly or completely illustrative of commitment itself. To my mind, a single promise implicates two different grounds of commitment: one of resolution, and one of fidelity. The commitments of resolution are related to the social practice of promising; the resolution in most (if not all) cases of promising is simply to

make a promise, or to participate in the social practice of promising, whatever that entails.[160] The second set of commitments refers exclusively to the content of the particular promise one has made and to which one is now bound. My promise to drive Bot to the airport binds me according to the general rules of promising – one must do what one promises – but *also* according to some specific source material – in this case, driving Bot to the airport. My resolution related to playing 'the promising game' cannot be completed unless and until I complete the specifics of my promise (or, if need be, make amends[161] for an inability to do so).

All the forementioned commitments of fidelity – promising, interpreting, analyzing, and explaining – are examples of an activity that further constrain some wider commitments of resolution. In translating,[162] it's not the text that primarily commits me to the source material, but the resolution I've made (or been given) to translate this text. The only legitimate means of completing this resolution is to limit myself to the content of the text – say, to the semantic meaning of the text in particular – and thus, I am committed to the text. But being committed *to* this text and being committed *by* this text can come apart. Both the explainer and the translator are *committed to* this text in the same way, but they are *committed by* it differently.

Those who translate a text for the purpose of presenting its substantive content are not necessarily working with the same source material as those who are aiming to create a literal

---

[160] We might disagree about whether the practice permits one to make a promise to engage in immoral behaviour, or whether one can promise to do something they never intend to or cannot do, but these issues are somewhat beside the point for me right now. These debates argue over what the associated commitments of the social practice of promising are (or should be) – what are the parameters of participating in the end of promising? Or, otherwise put, what are the rules of the promising game, and what counts as genuine or authentic participation? These rules will include not only the general moral thought that promises ought to be kept, but also associated moves such as what it takes to consider a promise kept, how to undo a promise that has been made, and perhaps how to make amends for a broken promise.

[161] As noted in the footnote above, I consider the social rules around making amends for a broken promise, and the rules around what must be done to escape a promise one has made, as themselves part of the promising game – just as rules around penalties are parts of any other game or sport.

[162] This is an example. I'm not staking a claim about the proper task of a translator, but providing an example about how, *given* some conception of "the proper task of translating," there are certain legitimate and illegitimate options.

translation of the same document. The former attempts to draw on the general sense of the work which may, for instance, benefit from using idioms in the translated language that do not exist in the original language, but which successfully capture the correct sense from the original author. The latter is concerned with the accuracy of the translation, rather than the understandability of the ideas in the translated language, and thus even an informative idiom would likely be thought an illegitimate translation.[163] Commitments of fidelity – and their precise arrangement in the commitment matrix – are thus highly informed by their grounding resolutions.

An apt example, given my use of constitutional precommitment, is that of judicial review and differing conceptions of how judges are bound when making judicial decisions. If your view is that constitutions entrench a specific conception of a given value[164] – such as what the framers originally intended by "equality" – then judges are significantly more constrained in what they may be allowed to decide *based on what is required by the constitution*. A judge under this constraint could certainly make a decision on a different conception of equality, but they could not *legitimately* claim that their decision is based in the constitution. Their commitments are more stringent, since not just any conception of equality is 'contained' in the constitution. Such a judge would act infideliously; and, to the extent that a judge's decision is validated or empowered by that constitutional commitment of fidelity, they would act *illegitimately*.

Alternatively, if your view is that a constitution entrenches the value, but not any given conception of it, then judges have a wider range of legitimate decisions open to them.[165] A decision relying on any conception of equality – insofar as it is, indeed, a conception of equality

---

[163] A more concrete example: those familiar with general discussions of constitutional interpretation will know the differences between interpretation based on original framer's intention and textualism. Even if the source material is, in an important sense, the same source, the text itself and a framer's intended meaning for the text can diverge. As such, it is important to consider these kinds of variants.

[164] This category of views, generally speaking, we might say are those of the "originalist" flavour.

[165] This category of views, generally speaking, we might say are those of the 'non-originalist' or 'living tree' flavour.

– is legitimate. There may need to be certain threshold for deciding when an idiosyncratic belief indeed becomes a 'conception,' but upon reaching that threshold, a judicial decision relying on that conception is legitimate. A judge under this model, then, may have significantly more means of legitimately completing their constitutional commitment (which may not be one of fidelity at all).

The main problem area for commitments of fidelity, outside of those associated with resolutions themselves, relate to disagreement about the source material, as there can be significant disagreement about the content of the source material. In promising, for instance, is the content of the promise determined by the promiser's intentions in making the promise, or the promisee's expectations, or by a disinterested third-party's interpretation of the utterance? What one may consider to be a peripheral detail another might consider a core feature, and I think that to the extent such disagreement amounts to disagreement about what the commitment itself really is or entails, it inevitably results in disagreement about how to assess some attempt at completing commitments of fidelity – in other words, disagreement about what is legitimate to do or expect.

It's also possible that tentative agreement on the source material is also plagued by differences of the grounding resolution, which may also result in disagreement about what may be legitimately done. Consider, again, investigations into the same source material that are grounded by two different resolutions: explaining the author's ideas and providing a word-for-word translation of the text. Two persons working on the same text but with different grounding resolutions may disapprove of how the other uses the source. There is enough similarity in their work that they can sensibly communicate with one another, but enough difference that

communication between them can break down if it is not remembered that they are pursuing different ends.

Finally, we can also have disagreements about whether and when a source material has changed, such as whether an amendment to a constitution constituted a mere addition to the former source material, or whether it is a whole new constitution with new framers. This especially matters if the source material is intended to be understood wholistically; a minor addition may change the full contextual meaning of what came before, such that previous components need to be reinterpreted in light of the new additions.[166] This may, of course, be yet another arena where the answers and conditions vary from field to field or are more or less stringent depending on the stakes associated. Anytime we're working with some 'source' or 'original,' there will to be room to consider and disagree about the metaphysics and identity of the thing.

Deference

The final type of commitment I will consider here is commitment of deference. A commitment of deference is one in which I am bound to some course of action because of the judgment and/or discretion of another, where that other may be a person or a decision-making process. These are the kinds of actions that are done and justified on the grounds that "X said so" or "this is how it's done." One's first thought, I think, might be that these types of commitments are characterised by law, or by parents' commanding of children, but commitments of deference needn't reach the level of such coercion or imposition.

[166] I'm reminded (perhaps unduly) of Ronald Dworkin's theory of constructive interpretation. Dworkin insists on the normative arm of this theory including finding that interpretation that puts the system of rules "as a whole" in its best light. This seems necessary on the assumption that additions do change (even subtly) that which came before. We might say that for Dworkin, the grounding resolution of constitutional interpretation is something like having a morally good constitution, and the source material (as he clearly points out) is the set of rules and principles of political morality.

Deference is complicated as, like fidelity, it also comes coupled with some grounding resolution – the reason for which deference is needed. If the need for a practice like precommitment arises, as Elster said, from rational agents "being weak and knowing it,"[167] the need for a practice like deference is, I would say, from being *ignorant* and knowing it. We may resolve to do something but realise we know quite little about the ends we seek. The resolution could be toward an end that is itself unfamiliar, or an end for which the means of completion are unclear: I may resolve to upgrade my computer but not know if it's possible or where to start; or I may know what upgrades I want, but not know which techniques are best to achieve them; or I may know both what to do and how to do it, but not know which products on the market are the best for what I want. In each case, there is a gap in my knowledge, and thus I seek out others opinions: I may take my computer to a shop to have someone else do the upgrades for me; or I may consult with someone knowledgeable about how to best complete the upgrades I seek; or I may seek out product reviewers to get a sense of what might be best for the job.

We don't only defer to the judgements of discrete persons, but also to processes. Science operates with deference to the scientific method – the process by which we think we will get data of a particular epistemic flavour. Science, we might say, operates according to the resolution of seeking truth – or perhaps, of minimising the chance of epistemic error.[168] Given this end, we've generated a process that, when followed, is intended (or believed) to output information that can be appropriately called knowledge (given what we consider knowledge to be or be like). Law too includes deference to process. Legislation requires following certain kinds of processes, thereby permitting a distinction between rules which have been correctly enacted (and are thus valid law)

[167] Elster, *Ulysses and the Sirens*, 36.
[168] Again, these are examples. I'm open to other ideas about what science's resolution is. Epistemology, I'm sure, is full of them.

and rules which do not pass muster as law (and therefore lack the authority or power owed to valid law). Trial courts may follow certain procedures of evidence or legal tests to ensure that legal processes achieve a particular kind of outcome (often relating to considerations of fairness or justice).[169] And don't forget our moral theories; the aim of comprehensive moral theories is to outline a process by which one will find (or is more likely to find) the "right answer" to a moral question, contingent upon what is thought to be the right moral resolution: for Kant's Categorical Imperative, the right resolution is to act rationally; for Mill's Utilitarianism, the right resolution is to act in a way that maximises net utility.[170]

The grounding resolution is what explains deference. When we defer to another, we do what they say *because they said so;* but their saying so only matters because of the resolution which grounds the need for deference in the first place. I boil my eggs as Gordon Ramsay says I should *because Gordon Ramsay said it,* but I only care that *he* says it because of my resolution to make the best soft-boiled eggs, and my judgment (however attained) that Gordon Ramsay is the authority who will get me there.[171]

The best-case scenario (from a deliberative point of view) is like the Gordon Ramsay example above, where my deference is rationally connected to my grounding resolution. We defer in our culinary resolutions to chefs who have a good handle on the best culinary techniques. When we're in the market for a new computer or makeup products, we may defer to

[169] Section 1 (the 'reasonable limitations' clause) of the Canadian Charter of Rights and Freedoms, for instance, requires that limits on charter rights be 'demonstrably justified.' As such, we implement the Oakes test to determine whether a purported constitutional limit is demonstrably justified: an infringement that passes the test is considered to have reasonable justification, and thus pass section 1.

[170] These are very superficial claims. I'm certainly open to discussion about what the grounding resolutions of different moral theories are. I mean only to demonstrate an example of what the different grounding resolutions might look like (beyond "do the right thing," which is not informative).

[171] My use of authority here is reminiscent of Raz's service conception, in that a person is authoritative if obeying their dictates – deferring – is more likely to get me to 'the right answer' than I would be if I tried to figure out the right answer myself. However, what Raz doesn't seem to discuss (that I've seen – though he's written a lot, and I've read but a little) is the role of any grounding resolution that orients the criteria for the 'right answer.'

the opinions of product reviewers who we think have a handle on what counts as a good brand or product. And we might defer to political leaders and judges who we think make good judgments about what's the best political or legal decision.[172] In other words, the best-case scenario for commitments of deference is, I think, when we defer to a person (or, perhaps, process) who is rationally connected, epistemically best suited, most authoritative, or most informed on the domain in which our resolution belongs. But of course, none of these rational connections are obligatory. We may defer to a makeup artist's culinary techniques, or a chef's make-up techniques. We may defer to a tv personality's medical opinions. We may defer to a non-descript celebrity on matters of international relations. This problem (and I do consider it a problem) can arise in a few different ways.

First, my lack of familiarity with a domain easily accompanies my failure to appreciate the epistemic boundaries *between* domains, and thus I may believe that one who is knowledgeable in one field or subfield must be knowledgeable in another. We might think, for instance, that our tax lawyer is surely also knowledgeable in criminal law, or that the artist who made my favourite ceramic pot will also be skilled working with yarn. There certainly are those – the 'jack of all trades' – who have a wide range of technical skill or theoretical knowledge of a general field. But, especially as specific domains and disciplines become more technical and nuanced, there's little good reason to think that *any* lawyer can answer your legal question, or that *any* artist can advise you with your artistic medium. The added difficulty, of course, is that it is sometimes necessary to belong to a field to know where these epistemic boundaries are. Some may have some minimal familiarity with things outside of their immediate skillset, like the professional flutist who has a basic sense of how a trumpet works, and do not attempt to claim

[172] Whatever the metrics of 'good judgment' or 'best decision' might be.

any expertise or authority.[173] But to those not familiar with embouchures or concert pitch, it is easy to think "a musician is a musician", thus the knowledge of one must be as good as any other.

Second, though we might recognize our own ignorance, we can fail to recognise it in others, especially if that other has marginally greater knowledge than we do. This is, of course, especially true if they lie about their knowledge or expertise, but this needn't be malevolent to be a problem. Some simply don't know what they don't know, and if asked about an issue or a process will give what they *think* to be true information. Others may be unaware that they're being asked for professional or authoritative judgments rather than simply their cursory, ill-formed or *prima facie* opinions. In other cases still, perceived success in a domain may be attributed to the person having some skill or expertise rather than luck or circumstance (or a misattribution of success altogether). Thus, celebrity heartthrobs might be asked for relationship advice, and wealthy emerald mine heirs who buy their way onto some successful company's board of directors might be asked for advice on starting a business.[174]

Finally, familiarity. We sometimes ask people their opinions on things they ultimately know nothing about because they, as a person or personality, are familiar to us. Perhaps we trust their judgment in other domains and thus will likely trust their judgment in the current one, or we feel the need to know if their views align with ours.[175] Thus, we ask the movie star for their

---

[173] The alternate is true as well. One may ignore saxophone tips for playing saxophone from the flautist because "they play flute; what do they know about the saxophone," but not realise that the finger positions are almost identical, ignoring register.

[174] Any resemblance to real persons, living or dead, is purely coincidental.

[175] This is an especially strange phenomenon with internet personalities, talk show hosts, and other celebrities who feel the need to or are pressured to announce their views on politics, global events, social movements, and the like. This is not to say such people shouldn't have opinions on these issues, but more that the expectation *that* they both have and promulgate an informed view is, to me, quite odd. I suspect it has a great deal to do with important issues becoming sources of entertainment with their respective fandoms. The hope is either that the beloved icon is on 'our side,' or that we can get a juicy tidbit for the tabloids. But this is an issue for another time.

position on some international crisis, or the internet personality their view on the national budget. This is not to say that movie stars and internet personalities shouldn't have opinions on international crises and politics, but rather that there is little reason to *expect* their views to be informative or authoritative.

These examples only encompass issues that arise under conditions of (what could be called) 'good faith' deference; where one who defers is really trying their best to ensure that deference is rationally connected to the grounding resolutions. But we may also simply be careless or lazy, and not want to be responsible for making decisions for ourselves; I might defer to the first opinion I come across because I can't be bothered to dedicate any time to this issue; I may defer to a powerful authority to avoid responsibility or criticism in case forging my own path proves mistaken; and I may defer only to those figureheads whose stances align with mine to avoid accepting the possibility that my position could yet be mistaken. These are still cases of deference,[176] and so long as my resolution in each case is just to avoid responsibility, to blend in, or to retrench what I already think, each is unfortunately legitimate.

Deference, as I see it, is not mere obedience. Under commitments of deference, the grounding resolution must be kept at the centre, as it is the reason for deference in the first place and thus the measure of legitimacy for one's commitment of deference. When deferring results in breaking the resolution, then deference is illegitimate. This can be highly complicated: what might appear to be illegitimate deference may in fact highlight our ignorance of the nuance of our grounding resolution, or a misunderstanding of our resolution, or an error in our following of the process. What might seem to be, for instance, an incorrect scientific finding could be a result of a poor research plan, a faulty experiment, or an insufficiently nuanced understanding of the

[176] As it should be clear, I'm not much of a purist of concepts or categories; abuse of a concept, to me, still counts as a use of the concept – albeit a bad one.

hypothesis. If deferring to the person or the process gets us closer to completing our resolution more often than not, then we might think those exceptional cases are simply exceptions or errors. A certain number of errors may be permissible (or inevitable), and those parameters would likely be determined by the person deferring (in the case of an individual) or by some recognised body (such as an 'academy of scientists'). [177]

But it's also important not to forget the grounding resolution, and to have a threshold for how much one will tolerate breaking that resolution. In early modern political philosophy, for instance, many of the works begin with a discussion of the state's purpose, and later build a theory of civil disobedience on the basis of those purposes.[178] In natural law theory broadly, as we know, an unjust law is no law at all and is not owed deference.[179] Deference is thus a skill, not only because one must know when to defer, but also because one must know when not to.

Forms of Commitment Compared

At this point, one might wonder what the difference is between commitments of fidelity and commitments of deference. On one hand, we might think that they're the same: when you're deferring to another, we might say that you're being fidelious to their command or judgment. In promising, which is a commitment of fidelity, the source material is the communicated promise – the utterance or statement of what one will do. In deference, we might be inclined to think that

[177] Margins of error in science research constitute exactly the kind of agreed-upon practice that I have in mind here. It's intended to serve as a sort of acknowledgement of the possibility of error and set a clear and rigid parameter for when such possibilities are too high to be reasonably acceptable.

[178] For Locke, "Where-ever law ends, tyranny begins, if the law be transgressed to another's harm; and whosoever in authority exceeds the power given to him by the law… ceases in that to be a magistrate; and, acting without authority, may be opposed." (Locke, *Two Treatises of Government,* Ch XVIII §.202). For Hobbes, "[t]he obligation of Subjects to their Sovereign, is understood to last as long, and not longer, than the power lasteth, by which he is able to protect them… The Soveraignty is the Soule of the Common-wealth; which once departed from the Body, the member does no more receive their motion from it." (Hobbes, Ch. XVII, §114). For both, when the power that justifies the role of the state is used in any way other than in pursuit of that role, the state acts illegitimately and is, importantly, not *owed* deference.

[179] For some, like Aquinas, there may be reasons beyond the rule's legal validity that may make obedience prudent, but the general sense remains. As a law, such a rule would not be owed deference.

the command or directive of the authority is some source material to which we're bound. In a sense, I can see there's some truth to this. To defer to another *is* to do or think as they say we must do or think. Nevertheless, I do think these different types of commitment look and function somewhat differently.

First, commitments of fidelity and commitments of deference put different emphasis on *the process* vs *the outcome*. Under commitments of fidelity, what you see is what you (hope to) get; you know upfront what the end result should look like insofar as you know what the source material is. When you're trying to complete something with high fidelity, like painting a lifelike portrait or accurately translating a text, there is a sense in which you already know, before you begin, what the outcome of that endeavour must look like if it is going to be fidelious. It may not be clear how to get to that end state, but what the end will look or sound like is apparent. In fact, *knowing* what the result is intended to be will shape the process one takes to get there; if I know that this statue is supposed to look like Aristotle, then I know there are restrictions on what the eyes and nose can look like, how aged the skin should look, what texture the hair should have, etc. I will know, from the outset, that I would fail if I attempted to make the eyes look any way other than the way Aristotle's eyes look.

Commitments of deference, on the other hand, may be better understood as emphasising the process rather than the result. I may seem to be splitting hairs here, since the point of deference is to complete a resolution of a certain kind, and thus to pursue some end. To the extent that I know my resolution is to make the best soft-boiled eggs or to make my laptop run faster, there's a non-trivial sense in which I know what the result should be. The difference, I think, is that while we know that deferring to this person or this process is supposed to get us to the end that we seek, we may have no knowledge of what the person or process will put out. It

could be that these are poor examples of deference, since in both cases the end is clear – great eggs and a fast laptop. But in cases of shifting goalposts[180] or dynamic resolutions, deference and fidelity seem (I think) to clearly come apart.

If I aim to be a responsible or ethical consumer but struggle to keep up with changes in corporate ownership, production and advertising laws, scientific discoveries and debunking, and so on, then it may be worthwhile (or even necessary) for me to defer to someone who is tasked with making these calls. It is perfectly in keeping with my resolution to be an ethical consumer, and in keeping with my deferring to their expertise, that I do seemingly contradictory things if they tell me to do contradictory things. My resolution – to be an ethical consumer – remains the same. What has changed is the objective determination of what is*,* in fact, the correct course of action in light of my resolution. If at t1 the judgment is to avoid Business A because of their unethical outsourcing, and at t2 the judgment is that Business A has changed their sourcing and is now ethically produced, I do not break my resolution or deference to act differently when the second directive is given.[181] This is because my commitment arises not from *what* they're telling me to do, but from the fact that *they,* as the authority on the matter, tell me to do it. The point is thus: under deference, one does not break this commitment when one performs an action that has been previously prohibited *if*, between the prohibition and the performance of the action, a permissive judgment has been given. Under fidelity, however, there is no such amendment; the same action was either going to lead to the final end or it wasn't. The conditions of completion haven't changed; we have either only corrected a previous error, or we are now in error.

---

[180] I do not mean this in the fallacious or manipulative sense, but in the sense more literally portrayed: I have a goal in mind, but what's required to get there is not firmly set in place. Thus, while I'm continuously aiming for the same goalpost, its placement relative to other ends and commitments may fluctuate. This will be especially relevant for commitments that rest on or are influenced by dynamic processes like weather, markets, etc. Aiming to save enough money for a down-payment is a stable goal in itself, but how much actually needs to be saved may vary over time. The amount needed is what I would say is the shifting goalpost.

[181] Again, this is only true if*,* in fact, the company *has* become ethical (whatever that means).

Another reason to think of deference and fidelity as different forms of commitment is that it seems possible for a commitment of one kind become a commitment of the other, and there's a significant and tangible change when this happens. Commitments of deference can turn into commitments of fidelity (and vice versa) potentially quite easily if the grounding resolutions are forgotten or changed. In cases of deference to persons, it is easy for charm and charisma to displace the grounding resolution that motivated deference in the first place. I may have resolved to live a more ethical life, and to that end found someone who may have seemed to be likely to guide me in this. However, if I lose sight of the grounding resolution, it is pretty easy for the status of their directives to shift: what I once considered and treated as being *conditional* judgments about how best to complete my resolution become, in themselves, a source material to which I feel I am bound even when they require me to behave in ways that would have broken the grounding resolution. My grounding resolution, at some point, has shifted from being a moral agent to being an obedient follower.

Importantly, this kind of shifting resolution needn't be nefarious or deceitful on the part of the 'authority;' this phenomenon is obviously possible through intentional maleficence.[182] But I think this kind of shift is actually *most* likely to happen when the authority is properly doing their job; stability can give rise to lenience or complacency, as the lack of *required* oversight may undermine belief of a *need* for oversight. When there is never conflict between the decisions of the authority and the underlying resolutions, it is too easy for the grounding resolutions to be forgotten as being the basis of those decisions; when the decisions always toe the line, it can be

[182] I'm reminded of a few different books I've read that demonstrate this point. *Animal Farm* (Orwell) is one – what started as a principled, egalitarian post-revolution living arrangement slowly turned into almost exactly what had been the case pre-revolution, if not worse. I think it demonstrates the ease with which previously devoutly held beliefs and principles can become either eroded or taken for granted and slowly forgotten (especially across generations), such that the committing force shifts from a principle to a person (or in this case, a pig).

easy to forget that there is a line being met. It is possible that the existence of those resolutions is itself forgotten, and all that remains is the habitual obedience to the authority. Thus, should the good-faith authority make an illegitimate move (relative to the original grounding resolution), it may never be noticed. After repeated errors, regular old conceptual drift may carry the decisions away from the line and the grounding resolutions, long since forgotten.

This is not just an issue of deference to persons. Even in the case of deference to processes, a process may take on a life of its own, becoming worshipped at the expense of those grounding resolutions which have been forgotten. We may have once resolved to have a court structure that ensured certain moral ends and established a process which was intended to achieve those ends. However, if the process has taken on a status of its own independent of these ends, then the legitimacy of judicial outcomes is no longer determined by whether it completes those original resolutions, but about whether we have accurately obeyed the process, regardless of how the judicial outcomes relate to those original moral ends. The process itself becomes a source material, and commitments of deference have, in essence, been replaced with commitments of fidelity. Failure to meet these values no longer serves any purpose in assessing legitimacy, and criticism is limited to whether the court has honoured the source material *rather* than whether or not they've fulfilled the original resolution.

Final Comments

To review: I have argued that it is mistaken to think of commitment as referring to immediate, ground-level decisions or to those decisions that are based on what we think is worthwhile, important, or significant. Rather, commitment is a higher-level statement of a certain kind of relationship between some state of affairs: actions and consequences, series of events, conjunctive or disjunctive relations, and so on. To be committed is not to have chosen one thing over another, but to be constrained such that one *must* choose the one or the other. What we

value, think worthwhile, or invest ourselves in may contribute to explaining the choices we make, but those choices are not, themselves, the commitment. The commitment is what happens such that a choice is made (or not), and the resulting consequences or implications that come into play.

Commitments are 'dealt with' in three ways. First, we may complete (or, perhaps, discharge) our commitment. In cases of completion, commitments are satisfied; there is, perhaps, an end sought which we have achieved, or a promise made which we have fulfilled. While there is some overlap between what completes a commitment and what is legitimate, this is not a one-to-one relation: one action may complete one set of commitments and break another set of commitments, in which case it is illegitimate.

Second, as mentioned, we may break our commitments. When the commitment continues to stand, but we fail to complete it, the commitment may be said to be broken. Breaking may be active or passive, depending on the type of constraint we're under. If I promise to do something, and I sit idle and do nothing, then I have broken the commitment passively. If I aim to stop doing something, and I do that thing, then I have broken the commitment actively. Whether commitments can be broken may depend on the source of commitment, and breaking certain commitments often commits one to further actions (like apology, making amends, or facing punishment or judgment).

Finally, commitments may be undone (or dissolved). Unlike broken commitments, which occur when the commitment remains standing yet is not completed, commitments which are undone cease to exist. Commitments may be undone via *release (*prior to expected completion) or excuse (after expected completion). Undoing commitments typically includes a change in the

grounding conditions on which the commitment was originally made, and whether commitments may be undone may depend on their source.

Commitments have two sources, each of which bear on the kinds of interventions needed to overcome or adapt to their constraint, and which bear on how (or whether) those commitments may be changed. Natural commitments come from the world as it is, while social commitments are forged by human activity. Natural commitments are operative outside of human control, and thus must be overcome through manipulation or waited out. Social commitments are fully within (collective) human control, and may thus be created, changed, and undone in their entirety through (collective) human activity.

Lastly, commitments take different forms, which affect both their relationships with commitments of other kinds, and also considerations of legitimacy in our decision-making. Resolutions are the most common form of commitment, referring to commitments wherein the grounds take the form of an end to be pursued or avoided. Both remaining types of commitment operate in conjunction with commitments of resolution. Commitments of fidelity take as their secondary grounds some source material by which the primary grounding resolution is refined. Likewise, commitments of deference take as their secondary grounds the discretion of another, whether person or process, to whom we defer for the purposes of completing the primary grounding resolution.

What I have put forward is a framework by which I think many philosophical discussions can be reinterpreted. I have mentioned some debates, domains, and theorists throughout the chapter, intending to demonstrate how many of my ideas do already somewhat exist in different forms, and also to raise some questions about holes and oversights in their current forms. It would take the current project much too far off course to give a more thorough discussion of

what these questions and theories might look like if reformed according to this model, but I do think it's possible and I think it would be interesting and worthwhile.

For our current purposes, however, this analytical work on the concept of commitment is intended to highlight the issues in the constitutional precommitment debate, and how the agent-dependent condition retrenches a number of the issues. As noted in the discussions above, what might appear to be minor, superficial differences of belief or motivation really do fundamentally change our commitment networks, thereby changing what we may legitimately do. As such, what appear to be minor disagreements about, say, whether constitutional precommitment is legitimate relies on extensive competing commitment matrices, and cannot be resolved through simple debate about the 'real' concept of democracy, or 'legitimacy,' and so on.

In the next and final chapter, I'm going to return to Elster's two problems of constitutional precommitment that I discussed at the very beginning: that constitutions may not bind at all, and if they do, then they may only bind others. Using the tools I have developed to this point, I can more deeply explain how these issues represent a fundamental misunderstanding of commitment, as well as some suggestions about how they also represent a misunderstanding of constitutionalism. While I make no claim or effort to rewrite the principles of constitutionalism, I want to suggest some alternative ways of thinking about constitutionalism that may help the whole "is constitutional precommitment democratically legitimate" debate move from where it's currently stuck.

Chapter 8: So, What is Constitutional Precommitment?

I have, throughout this dissertation, introduced a few distinctions and concepts that are meant to form a structured response to the problem of constitutional precommitment. I want to now assemble and display that structure and demonstrate how the tools I've made can help us make sense of where the debate is getting stuck.

Review

The important point to keep in mind, which I hope I have sufficiently explained throughout the dissertation, is that under my use of the word "commitment," *people* are not committed (except metaphorically). Commitment is a relationship between concepts and conceptual categories: commitments refer to the ways in which conceptual networks must bear on our reasoning because of the constraints that the different parts impose on each other. Commitment describes not my relationship to some choice that I've made, but instead how the choice I've made relates to other possible choices.

Suppose there exist two concepts, *X* and *Y,* where *if X then not Y*, and *if Y then not X*. I may freely choose X OR Y, but I may not freely choose both at once: if I choose X, then I choose -Y; if I choose Y, then I choose -X. Depending on the precise conception of commitment we're using, different statements may be said of what 'my commitment' is.

In Chapter 4, we considered both a 'folk' conception of commitment, which operates under one understanding of the relationship between X and Y above. On this ordinary language use of 'commitment,' if I choose *X*, then I would be said to be *committed to X*. A statement of my commitment, then, is a statement about what I value or think worthwhile. But, as discussed in Chapter 4, this folk understanding seemed to miss those uses of commitment that describe my being bound to things that, while not themselves chosen, arise because of what I've chosen: it's

not uncommon, for instance, to hear of someone being *committed* to a conclusion because of some different, preexisting belief, even though they have not stated any belief in the conclusion itself.

For this reason, we moved on to consider the academic conception of commitment instead, as explained through Sam Shpall. This conception treats commitment as the connection between that initial choice of *X* and the resulting consequence of that choice on my rational or moral relation to *Y*: while I *chose X*, I am in fact committed to -*Y,* and if I choose *Y*, then I am committed to -*X*. While the choices I've made serve as the *grounds* of my subsequent commitment, the commitment exists because of the normative relation that exists between *X* and *Y* which, as Shpall rightly points out, exists independently of my attitudes: my choice *commits* me, but it is not what I'm *committed to.*

Though I like many of the changes that the academic conception of commitment makes to the folk conception, I argued in Chapter 4 and 5 that it nevertheless still fails to hit the mark. The difference between these existing conceptions of commitment and the framework I put forward is in the role of agency as the *source* of commitment. The folk and academic conceptions presuppose the agent-dependent condition, which appoints acts of agency as the *creator* of commitment: commitments exist where agency has been exercised and cannot exist but for this fact. I've said a lot about the agent-dependent condition, but my general issues (in summary) are as follows:

1. It is assumed that exercises of agency are needed to explain how different people's commitments can differ so widely, but this begs the question: there only appears to be significant variation in commitments if we assume that people are only committed when they exercise agency. When we decentralise agency as the source of commitment, then there is relatively little difference in people's commitments that is left to be explained.
2. Choices are always pre-constrained. Appointing the choice one makes as the source of subsequent commitment ignores the fact that commitment also existed on the agent at the point of deciding, since the possible choices are themselves limited by the conditions

in which one finds oneself. For what variation there does exist, differences in these pre-constrained background conditions can go a significant way to fully explaining variation – and these differences are not necessarily a result of one's choices.

3. It is assumed that agency is how we differentiate between the potentially illegitimate commitment and the always strictly legitimate requirements. But evaluative concepts like requirement are nothing more than refined conceptual commitments: they are useful shorthand words that we use to describe certain conceptual bundles. They are subordinate to commitment, not apart from commitment.

On my model, commitments do not exist because I, personally, have done something to bring them about. Rather, commitments exist as a matter of natural or social 'fact,' independent of me and my individual activities.

For my framework of commitment, then, the statements of commitment from both the folk and academic conception above are incorrect. My commitment is not to that which I've chosen, nor is it to that which necessarily follows from what I've chosen. On my model, we might say that *I* have no commitment in any way but metaphorically. Instead, it is X and Y that are committed: the commitment that exists is ((X AND NOT Y) OR (Y AND NOT X)). When *I* make a choice – such as the choice of X – I metaphorically place myself into the category of Xs, which is committed to Y in such a way that my X-ness necessarily means I cannot have Y-ness.

This metaphorical language can make sense of both the folk and academic senses. The folk sense of valuing and investment certainly aligns with the idea of one choosing to metaphorically embody that which I choose: in choosing X, I conceptually arrange myself *as though I am X*, and I am thus committed by my metaphorical X-ness just as the category of X itself. The academic sense of consequences also follows: in conceptually arranging myself as though I am an X, I metaphorically occupy the very same conceptual space as X itself, which means I stand in the same relation to other concepts as X does.

But if the choices I make or my statements of value are not the sources of my commitment, then where does the normative pressure of commitment come from? This question

occupied us in Chapter 5, where I explained the two-fold distinction. This first distinction dealt with the sources of commitment and was intended to motivate us to think about the ways we are constrained in our reasoning and where those constraints come from. The sources of commitment refer to the basic *why* of our commitments: why do we eat, why do we sleep, why is the speed limit 80 km/h, why may I not take my neighbour's belongings, and so on.

As noted in Chapter 5, social commitments are *social* because they are created by human conceptual activity and would not exist *but for* human conceptual activity. Importantly, while social commitments are created by people, they are created on a collective level: *individual* habits are not 'social' commitments. Instead, I have in mind the larger-scale social games and processes like law, culture, social stratification and, perhaps, associated hierarchies or castes, and so on. They needn't be universal, but they cannot be individual; fringe group norms count, but an individual's habits do not.[183] Legal categories, like *person*, *defendant*, *citizen*, and so on, are categories that exist because of collective human practices, and would *not* exist but for the fact that *they are treated as existing*. As such, the commitments that we have in virtue of our belonging to one of these categories are social commitments.

Natural commitments, on the other hand, are those we have because of categories that exist whether human conceptual activity recognises them or not. Here we can think easily of biological categories, like mammal, or female, or infant: while the full contours of these concepts are obviously interpreted through (and thus at least partially constituted by) human conceptual

[183] Individual habits can certainly have some kind of normative significance – it is potentially meaningful when, for instance, someone suddenly begins acting differently from how they usually do. But this meaningfulness is not something I currently think is 'commitment.' Alternatively, one relevant way in which commitment may be at play on the level of individual action might be that there is an expectation (whether natural or social) that people's behaviour is more or less regular and consistent.

activity, we yet can appreciate the difference between a category like "mammal" or "carbon-based life form" and a category like "teacher" or "criminal."

The sources of commitment are not useful because they tell us what our commitments are; rather, the value of the distinction is in how the different sources provide insight into the nature of the commitments and how they relate to our conceptual frameworks. The two-fold distinction, then, can only help us if we accurately understand what kinds of commitments we have, for it does provide us with some clarity on how much (collective) power the relevant group has to make our commitments otherwise: we cannot change what it means for humans beings that we are warm-blooded mammals;[184] we (collectively) *can* change what it means for someone to be a criminal.

Knowing the sources of our commitment can give us some insight into whether, should humanity decide to abandon certain categories, the related commitments would continue to exist or not. We might, for instance, collectively decide to abandon altogether conceptions of racial, gender, or sex category, but some commitments related to these categories (such as physical development, propensity for certain illnesses, inherent sun protection etc.) will nevertheless remain intact. Depending on the situation, this may change the relative ease with which our natural and social commitments can be more clearly seen; what previously may have been thought to be a condition of one's race or gender may more clearly be seen to be related to social circumstance or environment, and what was previously believed to be culturally determined may in fact have been biological.[185]

---

[184] Out of an abundance of caution, I wish to be very clear: we may revise our concepts and change our own understanding of what a mammal is. We cannot change the fact that, even if we are wrong about those facts that 'mammal' is used to describe, the facts will. Disconnecting the concept of 'mammal' from, for instance, 'warm-bloodedness' will not change biology, though it may certainly force a change how we understand the phenomenon in question.

[185] Like most anything else, these possibilities bring with them great power for progress, and equally great power for abuse. But the fact that such an understanding risks abuse does not, I think, mean it is therefore incorrect.

For the individual, though, the sources of commitment are perhaps irrelevant; for one's own decision-making, it may not matter whether one is committed as a matter of nature or as a matter of culture (unless, perhaps, one has desires for activism and change). What *does* matter is the recognition that whether I am committed is not up to me. What *is* up to me is what I do with the commitments that I have.

It is not up to me whether or how or why I am committed. But, as discussed in Chapter 6, I *do* make different choices about how to confront that which I have no choice but to confront. This will obviously have downstream effects on my commitment matrix: as I make choice after choice, the network of commitments that actively bear on me will change my commitment landscape because of how the implicated concepts relate to one another. As I hope to have demonstrated in the Beet and Bot example from Chapter 6, one choice may limit some future choices, while another choice might expand the possibility for future choices: Beet's decision to farm instead of hunt means that his community's mobility is constrained, while Bot's decisions to hunt rather than farm means he is forced to uproot the community regularly. In this way, individual agency can bear on how one is committed. But it is not agency itself which determines whether one is committed: that exists outside of any one person's control.

As discussed in Chapter 6, the choices we make put us on one of three paths relative to the commitments in question: completing a commitment, breaking a commitment, or undoing a commitment. If I'm endeavouring to do something then there will be certain possible actions I take that will put me on the path toward accomplishing that end, some choices that will put me on the path to failing that end, and perhaps still other choices that take me off the path altogether.[186] If I'm resolving to be a lawyer, then going to law school would be one decision

[186] There are also choices I can make that have absolutely no bearing on the end at all; those are not relevant for this discussion, but I do want to make clear that not *all* actions or choices are a part of this decision-making process.

that puts me on the path to completion. Leaving law school would be one decision that puts me on the path toward breaking that resolution, unless I've decided I no longer wish to be a lawyer (in which case my resolution has simply been undone).

These decision-making processes operate on both the individual and the collective level. Since social commitments are commitments that exist because we, the people, treat them as existing, we can, in a sense, make certain choices about how some ends will be accomplished. Should we decide that we want those who practice law to be particularly knowledgeable about the contents of law and legal process, we may create systems to pick those people out, such as creating a standardised curriculum or requiring post-hoc standardised testing for licensing. But these, obviously, will only complete the end if they aid in understanding law and legal process. If one can satisfy the conditions of the curriculum without having the desired understanding of law or legal process, then we, as a group, have broken the commitment.

These processes are structured under particular forms of commitment, discussed in the previous chapter. If there is a particular end to be accomplished, then there will be a limited number of actions we may take which will be *legitimate* given the ends we hope to pursue for ourselves. The concept of legitimacy, from Chapter 6 again, refers to a category for sorting those paths which will get us to the end from those which will not, and is constituted by the conditions of the commitment matrix for which it served as a sorter. In other words, as I hope I have made clear, legitimacy's only purpose is to serve as a signal that a particular kind of sorting has been done: legitimacy (and the status of being legitimate) does not imbue upon the contents of that category any special moral status.

Commitment and Constitutionalism

With all these distinctions in place, we can now turn to how they'll come together for the issue of constitutional precommitment. Though the language of my framework here is unique, we will see that the ideas underpinning it, I think, are not.

Every natural law theory and most (if not all) political theories of note begin with some outline of what is taken to be the 'core values' or 'necessary goods' or 'ultimate ends' that law or government must protect or pursue if it's to be valid law or legitimate government. This is not simply some quirk of those writers which may be lopped off in subsequent work, but rather represents their keen understanding of what they're setting out to do. This first task, present in so many historical works on political and legal theory, is to my mind a 'taking stock' of our commitments, natural and social – as it is only after taking stock of our commitments that we can differentiate between what is legitimate and what is not. There is variation on what the different theories (and different theorists within the same theory) take our commitments to be. But there is a shared recognition of some basic facts of human life.

Across these different theories, there is a shared foundation that human beings simply cannot survive without society of some form. There can be no survival where there is no society. No human being comes into existence like Athena, sprung forth from her father's head fully grown. Given human vulnerability, the facts of human gestation and infancy, the harshness of climate and terrain, of poisonous flora and predatory fauna – in other words, given the natural commitments of living on *this* planet in *this* solar system as the kinds of beings we are – society of some kind is essential for survival.

Nor can there be society where there are no people. I think Hart's discussion of the 'minimum content of natural law' is a useful representation of the point just acknowledged: there are just certain conditions that must be met if we are going to be able to have a society.

> This simple thought has in fact very much to do with the characteristics both of laws and morals, and it can be disentangled from more disputable parts of the general teleological outlook in which the end or good for man appears as a specific way of life about which, in fact, men may profoundly disagree. Moreover, we can, in referring to survival, discard, as to metaphysical for modern minds, the notion that this is something antecedently fixed which men necessarily desire because it is their proper goal or end. Instead we may hold it to be mere contingent fact which could be otherwise, that in general men do desire to live, and that we may mean nothing more by calling survival a human goal or end than that men do desire it.
>
> [E]ven if we think of it in this common-sense way, survival still has special status in relation to human conduct and in our thought about it, which parallels the prominence and necessity ascribed to it in the orthodox formulations of Natural Law. For it is not merely that than an overwhelming majority of men wish to live, even at the cost of hideous misery, but that this is reflected in whole structures of our thought and language, in terms of which we describe the world and each other. We could not subtract the general wish to live and leave intact concepts like danger and safety, harm and benefit, need and function, disease and cure; for these are ways of simultaneously describing and appraising things by reference to the contribution they make to survival which is accepted as an aim. [187]

It is, as Hart says, a simple thought; if a human society is a collection of humans, then where there are no humans, there can be no human society. But where there is no 'society,' it is difficult to believe there can be humans (at least, for long). The resolution to create society is biconditionally connected to a resolution of survival. It seems then that on some minimal level, the commitments of social living are the commitments of survival.

Many of the commitments of survival come from our biology – in other words, many of these commitments are natural. If we are left as infants alone in the woods with no one to care for our needs, we will simply die. Even as adults, if we do not have food or water, then we cannot survive. If we do not have shelter from the elements, or protection from the harms of other creatures (human or otherwise), we may die – either from the assault itself, or because of a resulting injury which prevents us from securing food or water. Even if, in an exceptional

[187] The simple thought referenced here is Hume's (and others) 'modest' view of the of the ends of human activity: "Human nature cannot by any means subsist without the association of individuals: and that association never could have place were no regard paid to the laws of equity and justice." Hart, *Concept of Law*, 191.

circumstance, a small family-unit of people can exist for some extended period isolated from all others, there are known risks to inbreeding, which affect the survival of the unit. Absolutely none of this is surprising, novel, or (I would hope) controversial. We know we are vulnerable.

It was a point Hart recognised in his minimal content of natural law; there must be some level of basic protections if we're going to get any kind of cooperative enterprise off the ground. If basic needs of survival are not met, then there will be no individuals around to participate in the cooperative enterprise. Most don't simply wait around to die if the necessary resources are threatened or unavailable: if given the choice, people will leave or revolt. Certainly, there's no clear threshold of goods that we require (or bads that we will tolerate), but we know that if the limit – whatever it is – is met, people leave, revolt, or die, in which case society is in danger of dissolution. A commitment of social living, then, is to figure out and create the conditions under which people no longer wish to leave, have no desire to revolt, and aren't dying off.

Many of our natural commitments are straightforwardly knowable: we know we can't allow our citizens to starve or freeze to death (or at least, not too many at once). What makes people revolt or leave is less predictable, but not unknowable. While unhappiness and injustice are motivators for change and revolt, people can also tolerate a significant amount of bullshit before the powder keg explodes. Hart is keenly aware of this point, hence his noting that the minimal contents of natural law are compatible with quite evil social arrangements.[188] But we also know and frequently see in history that unhappiness may only fester for so long before it finally reaches a boiling point.

There may be no knowing what will be the particular spark that ignites the flame – especially because that spark might vary from person to person – but we can certainly learn from

[188] Hart, *Concept of Law*, 193.

history the kinds of things that seem a part of that cycle: corruption of the laws and in their enforcement; real or threatened loss of access to resources; cultural conflict in the population itself; the list goes on. It's one thing for a society of people, as a whole, to experience the pangs of famine and drought; it's a very different thing for one segment of the population to throw lavish parties with food and drink while the rest wither in the streets. It is one thing to live under conditions of constant and *consistent* legal rigidity and totalitarianism, and quite another to live under a system of rigidity for some and leniency for others. We know that the perception of unequal suffering (whether real or imagined) gets people going; we know that there is, at some basic level, an intolerance to *unjustified* inequality.

Obviously, this is still complicated. We might not always notice inequality, especially if it happens not to affect us personally. If it *does* affect us, we try to be reasonable and do not, in ideal circumstances, take every perceived inequality to be a personal slight; we allow for the possibility of error. We might also come up with – or be indoctrinated by – many a story of how certain structural inequalities are justified: effort and dessert, achievement and inheritance, the metals of our souls,[189] our actions in our past lives or the actions of our ancestors. But while these conditions (and perhaps a general reluctance to 'rock the boat') can go quite far in explaining complacency and tolerance of what may (or ought to) be intolerable, they are not bulletproof: if and when inequalities like these are seen to be unjustified, there is a risk of social implosion.

For the reasons above, then, those previously mentioned writers typically go on to outline some system which is meant to avoid (or at least alleviate) these issues. They outline systems of rules that are meant to create a predictable background condition in which people live, to prevent

[189] This is reference to Plato's Myth of the Metals, one 'noble lie' from the Republic used to explain the class divisions in Plato's hypothetical society.

(as much as possible) the spook of unpredictability or unreliability. They create and enforce rights that are meant to serve as tools to protect those who fear harm, and to provide recourse to those who were not sufficiently protected. They outline stable mechanisms of government to create a recognisable chain of command, in hopes of avoiding confusion or ensuring accountability.

These things matter, *not* because they represent the will of the people, or because they are the tenets of constitutionalism, but because they are (part of) the absolute basic conditions for getting people to live with one another. It is irrelevant how society is organised: if you want to have a society, there are simply some conditions that must be met given the basics of what a society is, and the basics of what people are.[190]

It does not matter, then, if we all abandon the game of constitutional precommitment: should human government as a whole abandon the notion of constitutionalism and everything associated with it, it will still be true that 1) people cannot survive without society, and 2) society cannot exist without people. There may be a mistaken tendency to think that the things contained in a constitution are important *because* they are part of the constitution, but this is backwards: such things are in the constitution because they are important in themselves.

There is a reason constitutions contain bills and charters of rights, and it's not because governments graciously extend protections out of good will or hope for praise. It is out of recognition that the areas around which these protections exist are areas necessarily considered for the state's survival: to fail to recognise the significance of these values (and others) is to risk

[190] Importantly, I specify the 'basics' here, because we can certainly have more or less robust notions of society and people depending on the ends we're hoping to serve.

the dissolution of society.[191] Whether from the state's own history or from broader world history, there are certain lessons we have learned about the kinds of things – the kinds of injustices, harms, concerns, etc. – that are like petri dishes for social and political trouble. A wise state will put such concerns front and centre to make sure that, in the course of regular law and politics, those boundaries are not overstepped.

But importantly, constitutional commitments are not our commitments *because* we have put them in our constitution. It would not matter if a constitution failed to contain protections for persons or property, or if a constitution did not outline the proper chain of command. It would still be true that such things are important for the continued functioning of that society; it'll simply also be true that their poor constitution will do little to help their society remain standing. Having a poorly made constitution may do little to actively help the state in question avoid these struggles, (and could, perhaps, actively expedite them) but it seems true that having *no* way of keeping these commitments in the public consciousness will increase the possibility of things going wrong. There needn't be a constitution, but there must be something.

Sources of Constitutional Precommitment

I opened this dissertation describing a 'mistake' that Elster believes himself to have made surrounding constitutional precommitment. Let's return to that mistake.

Elster speculatively extended his theory of imperfect *individual* rationality and precommitment to the domain of constitutional law, where he considered how state constitutions (and some of their features, like periodic elections) could be considered examples of this kind of precommitment on a more *collective* level.[192] In time, Elster came to denounce this idea, partly

[191] In some cases, extending recognition to certain areas is reactive; after a particularly tumultuous social period or after a successful revolution, certain things may be added to constitutions as a kind of reassurance or rebranding. Nevertheless, my point remains the same: whether in the constitution or not, those areas were the source of conflict and threatened the stability of society.
[192] Elster, *Ulysses and the Sirens*, 90.

owing to the internal structure of his own theory, but also in part in response to the backlash such a suggestion received. There were two main problems with the constitutional precommitment view (as Elster discusses them): constitutions may only bind others, or they might not bind at all.

It makes sense to me that Elster would see parallels to his own theory of imperfect rationality in the discussions of constitutional constraint. His motivation, after all, was to describe how people can work around the problem of weak will to achieve some kind of imperfect rationality. The language of constitutional constraint, even predating Elster, tends to describe how constitutions control our passions, represent our 'better natures,' concentrate our rational will, and so on. The language in constitutional law recognises the very same weakness of will (albeit on a collective level) that Elster was trying to confront in *Ulysses and the Sirens.* As such, it is no surprise to me that he'd make these comparisons and draw these parallels.

The problem with drawing these parallels, as Elster himself says, is that society is not an 'individual writ large.'[193] There is a vertical division in the individual which does not exist in the collective; in the individual, one part is 'properly' in charge of long-term planning. There is a presumption of stable preferences across time. There is an assumption of some part of these preferences being, in a sense, more 'oneself' than those preferences that are fleeting or against one's better judgment. As such, the mechanism of individual self-binding offers a way of expanding one's agency despite constraint, since the constraints help me to live more consistently with my 'true' self. But society has "neither an ego nor an id,"[194] and so there is no part of society who has an inherent claim to represent the general interest of society more so than any other.

[193] Elster, *Ulysses Unbound*, 167.
[194] Elster, *Ulysses Unbound*, 168.

This binding-others problem in constitutional law is more commonly known as 'the democratic worry', and is described through arguments around the 'democratic legitimacy' of constitutional precommitment. The best we could hope for, we might say, is that constituents could choose people to make up the group who will serve as their representatives. That elected group, we could then say, will have been given a claim to represent the general interest of society, albeit not in the same way that an individual represents their own interests. Still, there is something to be said for those people who have been *chosen* by the rest of society to be responsible for making the decisions and purporting to represent the general interest of the group.

The problem is that constitutions, insofar as they pass from generation to generation, seem to undermine this vesting of power in our chosen people. Given the supremacy that constitutions have over subordinate law (and subordinate law makers), even those chosen few are limited by a set of constraints which, in many cases, constituents did not choose for themselves. They're bound by constitutional rules put in place decades or even *centuries* ago, and while amendment is possible, the kind of coordination needed to make such changes puts constitutional amendment somewhat outside the realm of feasible possibility, particularly where there are many different forms and ways of life. Thus, in the case of most constitutional states, the constitution represents (at best) the will of the chosen few from several generations ago and serves as an unchosen constraint on those who follow. Current constituents are not free to set the rules according to their own rational wills, and are instead at the mercy of the dead hand of the past. Constitutions, then, seem only to bind others.

As discussed in Chapter 1 and 2, the standard approach to the binding-others problem is through a democratic angle: the debate has turned into arguments about the concept of

democracy and what it requires. The critics of constitutional precommitment – those who are sympathetic to the binding-others problem – are worried that unchosen constitutional commitments seem antithetical to democratic self-governance: if 'we the people' have not chosen the contents of our constitution – which serves as the foundation for our government and laws – then 'we the people' are *not* self-governing. Instead, we're governed by whoever created the constitution which now binds us.

Advocates of constitutional precommitment, however, *also* argue from a democratic angle. In Chapter 1 and 2, I briefly discussed two general forms of argument that advocates give in response to the critics on this issue. The first form of argument I called the 'metaphysical argument' and defends constitutions as self-binding on the basis of a collective conception of 'self': it defends the idea of an atemporal, collective identity – the idea of 'a people' – by whom and for whom the precommitments have been chosen. The second line of argument is what I have called the 'democratic precondition,' which argues that constitutional precommitment can't be democratically illegitimate because democracy is impossible without the constraining power of constitutional precommitment.

What I really want to emphasise here – and what I think serves as so much of the hang-up in the constitutional precommitment debate as it currently stands – is that these two groups of people are *both* claiming to argue from *what democracy requires.* Both critics and advocates think that democracy is the best conceptual tool with which to solve the precommitment puzzle; neither camp intends to put forward an anti-democratic argument. But it is also clear that the two camps fundamentally disagree about what democracy is. Advocates of constitutional precommitment think that the critics neglect important substantive content that democracy presupposes, while the critics think that advocates are shoehorning their personal moral beliefs

into the democratic process. Much like the translator and the explainer from the previous chapter, it's not enough that the two groups are using the same language and can seemingly communicate with each other about their projects; they are very clearly using and engaging with these ideas in different ways, and seem to have different and incompatible resolutions. As such, if we try to continue along this line of argument, the constitutional precommitment debate will be unable to go anywhere unless and until the two camps somehow make progress on the issue of what democracy 'really' is.[195]

The value of Elster's so called 'mistaken' application of a theory of precommitment to constitutionalism is that it is, as far as I'm aware, a novel angle from which to map out the space. So many approach the problem from some other angle – Democracy, Good Government, Natural Law, and so on – but despite being part of the theory's name, the role of *commitment* in constitutional precommitment has been given glaringly little attention. If we can't make any progress on this issue from the perspective of democracy or 'good government,' then we need to try a different angle. Elster was indeed mistaken when he first tried to extend his view of imperfect rationality to constitutions, but this is not because constitutions do not or cannot commit us; rather, the error was the fact that he tried to extend a theory of indirect agency to a condition of non-agency.

If I were analysing constitutionalism through the lens of commitment, then, I would respond to the binding-others worry by highlighting that it presupposes the agent-dependent conception of commitment. One reason the binding-others problem is believed to be such a

[195] I'm not precluding the possibility that one side or the other (or both) is simply mistaken about what democracy 'really' is. But the burden on those who wish to make these claims includes weighty claims about the correctness conditions of democracy: beliefs about a correct conception presuppose certain attitudes about what our concepts should do for us. When it comes to concepts like democracy which are, to my mind, social constructions (or social kinds), these are necessarily normative attitudes, not descriptive. Either way, until there is somehow a singular understanding of democracy, this issue will persist.

strong objection is that it operates on the assumption that commitments are agent-dependent, and thus can only exist where they have been created by the agent. If this were the case, then the binding others problem would indeed be a serious issue for the coherence of constitutional precommitment. On the agent-dependent view, it would indeed be conceptually true that whatever has not been chosen for oneself *by* oneself cannot possibly be considered commitment. But if commitment is *not* agent-dependent, then this objection loses a good portion of its bite. It's not being chosen that makes a constraint a commitment; rather, it's a commitment because one cannot choose but to confront (or neglect) it.

There are, of course, remaining concerns about a lack of democratic agency. Even if one can be committed without choosing to be so, and subsequently have constitutional commitments, it still means that those constitutional commitments are undemocratic. To this worry, I say: fine. I'm willing to concede that under some conceptions of democracy, a constitution that contains unchosen constraints is indeed undemocratic. My point is only to emphasise that it's not "being chosen" (or, by extension, "being democratic") that makes us committed. In the interest of full disclosure, I think constitutions *are* undemocratic, as I think that's precisely the point of having a constitution serve as a check on political power (no matter its form). But I also don't think that the lack of democratic legitimacy is quite the problem that the more democratically minded take it to be.

To me, saying something is undemocratic is simply to say that it is democratically illegitimate. As we remember from Chapter 6, my use of 'legitimacy' is connected to a checklist against which a thing is assessed. In this case, when we say unchosen constitutional constraints are undemocratic, we are saying that there is a checklist of commitments – features, conditions, rules, concepts, etc. – related to the category of things that are called 'democratic,' and that

unchosen constitutional constraints do not satisfy the checklist. In other words, the conceptual structure of constitutionalism breaks some democratic commitments, resulting in a condition of illegitimacy.

With all this, I'm quite in agreement. If it is believed that choice and agency are core features of the democratic checklist, then it would indeed be true that conditions of non-choice and non-agency are democratically illegitimate. So, for both critics and advocates, it's true that a condition under which the demos has not chosen its commitments is indeed an illegitimate condition. Both groups think the concept of democracy includes something about choice and agency, which means that the legitimacy conditions of democracy require choice and agency; they simply disagree on whether that legitimacy condition has been met. But as I also suggested in Chapter 6, *illegitimacy* is not inherently bad; the fact that it's illegitimate does not necessarily mean it is therefore wrong, immoral, or ought not be done.

This point is not one that critics and advocates would (or even could) agree with, since it cuts directly against their legitimacy conditions. For both groups, the concept of democracy builds in certain normative claims about a democratic decision. The critics, with the more majoritarian conception, will attribute a kind of moral legitimacy to the outcome on the basis of its having been consented to in the appropriate way: it furthers liberty, or self-governance, and is therefore *good*. The advocates, on the other hand, attribute to the concept of democracy such a significant moral threshold such that a democratic decision *by its very nature* is good – if it were not, it would not be democratic. For both groups then, it is indeed true to a certain extent that democratic illegitimacy is morally bad: some threshold of moral goodness is presupposed by the democratic conditions of legitimacy.

But I am not convinced by any inherent moral goodness in democracy. On the one hand, I agree that the advocates' conception of democracy presupposes far too much moral content to be a meaningful or useful operative concept; so much is required that the concept (to me) offers little descriptive value. On the other hand, though, regardless of whatever legitimating value exercises of liberty might offer, I am not convinced of any moral goodness offered by a more majoritarian conception as used by the critics. Liberty is one value among many; it oughtn't be the end of moral consideration.

If I may, consider the following. In Chapter 6, I suggested that the fact that we do not choose our commitments may arguably be what makes agency so valuable since, though we can communicate with each other about our values, desires, ills, and so on, no other person has the very same access to our minds and decision-making processes as we do. While we share a significant number of our commitments, it's also true that differences in time, space, capacity, resources, etc., mean that personalised reasoning and judgment is often required for us to be able to make *legitimate* decisions (if and when legitimate options are available).

If we are trying, as agents, to navigate these numerous physical and psychological constraints, then *we* are best able to do so; only the individual agent themselves has access to their inner world. We can ask others for advice and provide them (to the best of our ability) with details we think are necessary for them to help us with our decision. We may also be given unsolicited advice by those who think they know our inner world, or who assume our inner worlds are similar. But in each case, the individual themselves has the best access to the inner world. This, again, is not saying that the individual will always make the best decision for them; one can be a bad decision maker even about what is best for oneself. This is only to say that the individual themselves is either the most *likely* person to make the best decision or is the *only*

person who *could* purposefully make the right decision – and if they cannot, then no one else ever could.

If we're trying to ensure that society can exist and endure, we need to make sure, as noted above, that people do not leave, revolt, or die. As a shorthand, then, we might say that an enduring society requires some level of 'happiness,' and that unchecked unhappiness will result in citizens leaving, revolting, or dying off. One way to make people very unhappy is by ignoring their wants or needs, either because you assume you know what is best for them (paternalism) or because you simply don't care about their wants and needs (tyranny). If we're trying to ensure that society stays together, then this sort of steamrolling of desires is one thing that complicates that end.[196] As noted above, if the people are (or think they are) incapable of securing their needs, then they will (eventually) leave, die, or revolt.

How might this relate to democracy? Quite simply, it may just be that the best way to figure out what people want is to *ask them.* If, as a leader, you have resolved to find the course of action that will minimize the risk of your citizens leaving or revolting, then asking them what they would want you to do seems (at least *prima facie*) helpful for completing that end. Democracy is little more[197] than a political decision-making structure that rests on the observation that *asking people* what should be done about things that affect them is useful for keeping society together. There are certain issues we simply must confront, and letting the group have a say may be just the thing to make sure that they do not leave.

---

[196] There are many theories as to the push-pull factors of human migration, but there seems to be (according to my naïve understanding) a clear trend that unhappiness and/or instability drive exodus.

[197] With this, I'd wager, all theorists of democracy would agree. Where they disagree are on the conditions under which one counts as "the people," the conditions under which one counts as having been asked what one wants, what it means to want something, and the conditions under which one has given an answer about what one wants.

In another arrangement, we could say that democracy is a commitment of deference (to the democratic process), the grounding resolution for which is to make political decisions that do not threaten the stability or existence of society. This understanding fits with and may explain the critics' conception of democracy: the legitimating power of liberty or agency rests in the fact that it is (more) likely to result in the individual agent's making the decision that works for them, given their particular commitment matrix. We follow this process – putting forward a decision to the people and allowing them to declare what they think the legitimate choice is – because we think it will get us results of a particular kind, such as avoiding or lessening the risk of mass death, exodus, or discontent (because, as we remember, those will threaten society).

Obviously, this is imperfect: people don't always have all the information to decide; had they been given different information, they may have made a different decision; and sometimes, people simply don't care and will opt for any option simply to be free of the responsibility to decide. In these conditions, deference to the democratic process may not be enough to complete the resolution of avoiding social instability; if the people make a decision on the basis of some basic information, and further information is made available after the fact, then the people may feel that they were deceived or manipulated. Even if they have all relevant information, the people may find that they did not fully understand the ramifications of their chosen outcome, and come to regret their decision after the fact, which may result in unhappiness. People may also choose hopefully, unaware of or perhaps willfully blind to the fact that they, personally, do not have the means to support the social or economic implications of their decision.[198] In short, for a variety of reasons, people can make bad decisions (even when we are trying, in good faith, to make the right ones).

[198] Consider, for instance, the 'temporarily embarrassed millionaire' who votes for tax cuts for the wealthy, even while they are personally dependent on the programs that those taxes support.

This is another commitment that we share: we are all capable of making bad decisions, no matter our metric of what constituted 'good decision-making.' If the dissolution of society can only be avoided with 'good decisions,' however determined, then how do we deal with the fact that people sometimes do not act toward the common good, or act with rational self-interest, or even act in good faith? As I said above, it may be true that only the person themselves could know what is legitimate for themselves, given their particular commitment matrix. While it may be true that no one else can know, in a sense, what is 'good' for us to choose, it is *also* true that we could yet know when some decisions will be very bad.

These commitments are, no doubt, why competing conceptions of democracy exist. As noted previously, if our checklists consider legitimate that which we think ought to be *il*legitimate, then we can change our checklists – our conditions of legitimacy – to change what subsequently 'counts' as legitimate. So, if it's a commitment of political decision-making that people can make bad political decisions, some think we should make it a condition of political legitimacy that decisions meet some minimal moral threshold (however that is determined); and if it's a problem that people may make different decisions with different information, some think we should not count one as having legitimately decided unless one is fully informed and cognitively unencumbered when deciding; and so on. These addendums and amendments to the critic's conception of democracy will get us something that looks more like the advocate's conception of democracy.

But importantly, the resolution at the bottom of all this is, presumably, to have an enduring society (for the yet further resolution of survival). As discussed in the previous chapter, this grounding resolution sets the parameters for the available courses of action relative to the end pursued: if society is our end, then our political arrangements *must* work toward completing

that end (if, of course, the point of political decision-making is indeed for society). If the particular political process takes on a life of its own, then it risks becoming removed from that grounding resolution; the narrow focus of ensuring that decisions have satisfied the particular conditions of the decision-making process may result in a condition where those decisions themselves cut against the grounding resolution. And if too many decisions are disappointing, unfair, or endlessly delayed, then the people will begin to leave, revolt, or die in the meantime.

Clearly, then, there are good reasons to have democratic political decision-making processes: at the very least, asking people what they want is inherently easier than trying to guess. But the fact that there are good reasons for it absolutely does not mean that there are no reasons to be wary of it; people may not know what they want or, as Elster so plainly demonstrated for us, may be conflicted in their preferences. As such, despite whatever good there might be in the democratic process, there is yet good in ensuring that it is limited.

This may seem that I have simply reinvented the democratic precondition argument from Chapters 1 and 2. But the important difference between the precondition argument as offered by the advocates and what I'm saying here is that I *agree* with the critics that constitutionalism conflicts with democracy. Where those who espouse the democratic precondition argument can be thought of as adding these conditions to their concept of democracy itself, my argument here is that those constraints exist *apart* from the concept of democracy. While they might argue that democracy requires a certain level of liberty and equality in order to get off the ground, and thus democracy presupposes some moral threshold, *I* argue that stable society may require certain levels of liberty and equality to subsist, and thus this moral threshold must serve as limits on *any* political arrangement. And thus, while it seems implied, according to the advocates, that an 'evil'

democracy could never be a 'true' democracy, I think such views are naive and risk lulling us into a false sense of security.

I am reminded of early positivist concerns about natural law theory and the insistence on the moral quality of law.

> [T]he time might come in any society when the law's commands were so evil that the question of resistance had to be faced, and it was then essential that the issues at stake at this point should neither be oversimplified no obscured. Yet, this was precisely what the confusion between law and morality had done…There are therefore two dangers between which insistence on this distinction [between the existence of a law and its moral quality] will help us to steer: the danger that law and its authority may be dissolved in man's conceptions of what law ought to be and the danger that existing law may supplant morality as a final test of conduct and so escape criticism.[199]

As Hart discusses, the confusion between the procedural validity of a law and its moral quality risks complicating the process should the time come that laws require behaviour so evil as to not be worth obeying. This, again, is ultimately concerned with the observation that some minimal conditions must be met if people are going to live together; and if a law requires us to do something that would threaten the ability for society to function, or risk social stability and security (however determined), then we need the resources to be able to see the issue for what it is: the law is valid, and immoral, and must be changed lest it threaten the persistence of the state.

It is the grounding resolution of social living which informs the positivist concern about confusing law and morality; it is the grounding resolution of social living that also informs us of the confusions between a democratic decision and a good political decision. If abiding by the democratic process will create conditions under which the grounding resolution will be threatened, then we need to interrogate (or abandon) the democratic process. But just as is the case with Bentham's reactionary, who believes that "this is the law, therefore it is what it ought

[199] Hart, "Laws and Morals", 597-98.

to be," the fact that something is the result of the democratic process might also "stifle criticism at its birth." [200]

One of the benefits of recognising the conflict between constitutionalism and democratic self-government, I think, is that it forces us to pay attention to how the two domains interact. Both give us legitimacy conditions: certain conditions must be met as given by the constitution (and constitutionalism more generally), and certain others as given by the concept of democracy. It is not *bad* if the constitution is 'democratically illegitimate,' as that is precisely the point: if that which is meant to serve as a check on power never exists in tension with that power, it is questionable how much of a check it can really be. In as messy a domain as politics, one ought to expect conflict to arise at nearly every possible opportunity. A condition free of political conflict, far from something which ought to be reassuring, seems to me instead a condition worthy of scrutiny. If, to minimise opportunities for conflict or to seem accommodating, we define our constitution as abstractly and minimally as possible, then it does, indeed, seem that constitutions can be little more than parchment barriers.

The Constraint of Constitutional Commitment

This brings us nicely into our next problem, which was the argument that constitutions simply can't bind anyone at all. A constitution is, at best, a single piece of paper with some words written on it and, at worst, a set of abstract ideas hidden and dispersed across years of court decisions and social practices. A constitution is incapable of knowing itself, let alone of enforcing itself or speaking up for itself when its contents have been contravened. If we think that commitment is supposed to be some stringent constraint on our behaviour, then clearly that

[200] Hart, "Laws and Morals", 589.

which is little more than a 'parchment barrier' cannot serve as any meaningful source of commitment. But I think this is wrongheaded.

There are two general ways that I think the non-bindingness problem misconstrues the kind of constraint that commitment really is, and thus how it misconstrues the value of a constitution (and its constraining power): first, the non-bindingness problem understands commitment to be too physical or literal in its constraining power; and second, by failing to accommodate the possibility of a 'broken' commitment as such, it neglects the value of broken constitutional commitments.

In Chapter 3, I mentioned that the non-bindingness problem is (somewhat) a result of Elster's conception of precommitment, as his criteria for precommitment in *Ulysses and the Sirens* were quite stringent: in order to count as 'proper' precommitment for Elster, it must be the case that my precommitment at time t1 change the set of actions open to me a time t2 such that the action to be avoided is not available at t2. In other words, if the purpose of my precommitment is to avoid performing action X, and I remain able to perform action X, then I have not been precommitted.

There are at least two ways to interpret Elster's criteria that precommitment remove some action from the realm of possibility (or feasibility). On one interpretation, constraint is understood to be physical; the existence of the constitutional commitment against *X* must mean – to be effective precommitment – that doing *X* is fully removed from the realm of possible future action. Constitutions could never fulfill this condition, for the reasons noted just before. Everything that can be done constitutionally can be undone, and there can be no external mechanism that ensures constitutions are obeyed.

It seems obvious to me that this first interpretation is one that Elster ultimately follows, given his subsequent conclusions about the disanalogy between constitutional precommitment and individual precommitment. And, to his credit, the conclusions that constitutions do not bind at all *do* seem to follow from the premises on this interpretation of constraint. His criteria for precommitment require i) that one increase the likelihood of performing a certain action at a future time, and ii) that *some* options are removed from the set of potential future actions. From these criteria alone, it makes perfect sense to me that Elster rejects the view of constitutions as tools of precommitment *as he understands it.* It cuts against at least two of his five criteria of precommitment, and therefore can't be precommitment by definition.

It is, I think, forgivable to interpret this condition as requiring a *physical* constraint or creating a condition of *physical* impossibility, as many of the examples that Elster uses throughout *Ulysses and the Sirens* are indeed tangible and physically constraining in nature: blocking access to the thing to be avoided, setting in place causal mechanisms to ensure a certain outcome, and, of course, being physically tied to the mast. And, as evidenced by Waldron's concerns, this is indeed the interpretation that seems to have taken over.

But as also discussed in Chapter 3, Elster himself recognises that physical constraints may have minimal usefulness, (though he seems to have forgotten those ideas by the time of writing *Ulysses Unbound)*. Let's remind ourselves of how he thinks this through. Elster outlines the different methods of subverting weakness of will that he considers through his discussion as follows:

1. Manipulation of the feasible set
    a. Restricting the set of physically possible actions
    b. Changing the reward structure by public side bets
2. Manipulation of Character
    a. Strengthening will-power

b. Changing the preference structure

3. Manipulation of Information
    a. Changing the belief system
    b. Avoiding exposure to certain signals[201]

And, as discussion on these methods, he offers the following appraisals:

> Within the set of actions (1) that change the feasible set I distinguish between, say, (1a) the strategy of going for a long walk in the mountains so as to make cigarettes physically unavailable and (1b) the strategy of telling your friends that you will stop smoking so as to change the reward system. The (1b) latter strategy also induces a change in the feasible set, because the option 'Continue to smoke without any sarcastic comments' now becomes unavailable. Within (2) character-modifying actions I distinguish between (2a) the general strategy of strengthening the will-power and (2b) the more particular strategy of modifying some particular desire; the first permits you to climb higher uphill slopes while the second lowers the height of the slope that is to be climbed. Within the set of actions (3) that modify the information upon which further decision are taken I distinguish between (3a) the radical strategy of inducing new factual beliefs (including a strategy for inducing forgetfulness about the induction) and (3b) the moderate strategy of screening yourself from certain signals or cues in the environment.[202]

Again, we notice, in particular, that the method of changing the reward system (1b) is remarkably metaphysical.

> In most actual cases strategy (1a) is unfeasible or involves sacrifices too heavy, and would thus fail to satisfy criterion (iv) of II.2. The same, I think, holds for method (2a), which would often involve a form of overkill whose feasibility might do away with the problem itself... I submit that in most everyday cases the most efficient strategy involves a combination of methods (1b) and (2b). Realising that through a series of actions I can achieve a hexis from which the desired actions will flow naturally ('sans violence, sans art, sans argument'), and that each of these actions is outside the immediate reach of my willpower, I may precommit myself to them by changing the reward system.[203]

If we understand commitment (and precommitment) as creating the kind of physical impossibility that is typically presumed, then the non-bindingness problem is a slam dunk: constitutions are useless at binding fast our hands. Again, one of the major impediments to the coherence of constitutional precommitment is that everything which can be constitutionally done

---

[201] Elster, *Ulysses and the Sirens*, 103.
[202] Elster, *Ulysses and the Sirens*, 103-4.
[203] Elster, *Ulysses and the Sirens*, 104-105. As a reminder, criterion iv is "The resistance against carrying out the decision at t1 must be smaller than the resistance that would have opposed the carrying out of the decision at t2 had the decision at t1 not intervened."

can be *broken* (by outright ignoring the constitution) or *undone (*through formal amendment*)*, and thus nothing is ever really removed from the realm of possibility.[204]

And yet, it doesn't even seem to me that this first interpretation best fits Elster's original views. The second interpretation of Elster's criteria considers more these intangible constraints. On this interpretation, precommitment doesn't work by physically binding us like Ulysses to the mast, but by reorganising our rational (and thus logical) world by changing how categories relate to each other – through *commitment*. The constraint that a constitutional commitment provides is *conceptual*, not physical. Constitutions outline processes that, when followed, create "legislative power" or "valid law." They set out areas where that "legislative power" doesn't operate, or where it operates under rigid constraint. The purpose of constitutional precommitment is not to absolutely prevent the possibility that those under its jurisdiction fail to be 'properly' prevented from doing what they are not supposed to do, or 'properly' forced to do what they are expected to do. Rather, constitutions create a conceptual landscape wherein connections are forged between certain behaviours and outcomes, such that certain desired or desirable qualities (like validity, or authoritativeness) can *only* be acquired after following the process. Constitutions generate and arrange categories of conceptual organisation that then inform how we think about those categories.

In this regard, it is helpful to think of the practice of promising, insofar as constitutionalism and promising both appear to function at two different levels of analysis: the individual level (of the individual promise or constitution) and the general level (the rules of the game of promising or constitutionalism). There is, it seems to me, an undeniable change to an action after it has been made the subject of a promise: what might have had little to no normative

[204] Waldron, *Law and Disagreement*, 260.

value before the creation of a promise suddenly becomes a core normative case. For me to promise that I will not eat the last piece of cake in the fridge and then eat it is different than if I ate the last piece of cake in the fridge when no promise has been made: the *existence* of the promise, somehow,[205] changes the moral flavour of the action of eating the piece of cake.

Obviously, the fact that I have made a promise may do absolutely nothing to ensure that I do not do that which I've promised not to do. Just like constitutions, it seems, promises are little more than conceptual constraints. If I have some level of integrity, then I might care that I've made a promise, and may thus gain some additional reason or motivation to follow through. But if I don't particularly care about the promise I've made, a promise cannot force me to follow through. Nevertheless, something happened to me – or to my normative existence – after I promised not to eat the cake: unless the background conditions change, it is impossible for me to eat the cake without doing something wrong, since 'eating the slice of cake' and 'engaging in immorality' have become committed in my normative matrix. The kind of constraining power that a promise has, we might say, is intangible or conceptual: it doesn't prevent the action, but changes its normative character or quality.

There is a game of promising, governed by general rules that apply to individual instances of promising. Regardless of the particular details of what I have promised, I have some overarching source of commitment relative to *the social practice* of promising: what it means to make a promise, what kinds of promises may be made, the scope of a promise once made, the conditions under which I may be released, and the effects or implications of breaking the promise. Because this practice exists independently of me, *I* do not harm the practice by breaking

[205] I have no hard theory of promising here; I anticipate that my claims about the change in moral flavour are compatible with a number of different explanations or justifications. But it does seem to me that, insofar as someone thinks the act of promising is morally relevant, it changes the moral quality of the action.

my promise. Rather, my broken promise (and perhaps even undone promises) results in a condition where my actions or my identity take on a particular flavour relative to the general practice of promising: I come to be seen as flaky, manipulative, or untrustworthy.

The game of constitutionalism is the same. The problem at the core of the non-bindingness objection is that, because extraconstitutional action is always within the realm of possibility, constitutions can't really bind. But this misses the fact that the whole *point* of a constitution is to set up conditions and expectations for how a state will behave relative to *its own claims*. There is a fundamental difference between a state which claims to be bound by a constitution and a state which makes no such claim. While there might be certain other conditions according to which the non-constitutional state is judged (relative standards of ethics, international codes of human rights, etc.), the state which *claims* to be bound by a constitution opens itself up to interrogation about its behaviour in a way that the non-constitutional one does not. A state which *claims* to be bound to a constitution, yet seemingly goes against all that it claims to be bound by, will be a state which is (or ought to be) mistrusted by *other states.* A constitution, in this sense, is no different than a promise; the fact that a promise can always be broken does not mean that there is no constraint in the existence of a promise.

The same state action takes on a different normative visage after it has become the subject of a constitutional constraint in the same way that the act of eating the slice of cake changes after it becomes the subject of a promise. It does not matter that the constitution cannot physically prevent the state from engaging in that action: what matters is that henceforth, that action is interpreted relative to the existence of the constitutional constraint. A constitution which prohibits action X does nothing to physically prevent us from engaging in action X. What is removed from the set of future options is the ability of truthfully uttering (or behaviourally

demonstrating) both "This is our constitution" and "We do not abide by this constitution," which becomes akin to denying ever having any intention to perform an action that one has promised to do; such an activity is inherently illegitimate. Of course, we're familiar with the concept of a *false* promise: I can certainly promise to do something that I have, in fact, no intention of doing (perhaps to induce some desired behaviour in my promisee), but such behaviour would typically be considered an abuse of the institution of promising (if a promise at all) and will reflect poorly on me as an agent. The fact that an agent can break a promise does not detract from the normative power of promising.

This, then, brings us nicely to the second way in which the concerns of the non-bindingness problem misconstrue the constraining power of constitutions. The fact that constitutions include (and, I'd say, *must* include) an amendment formula is taken to be a reason for thinking that they cannot be tools of precommitment, as what is done may always be *undone.* But the fact that constitutional constraints are undone (through their amendment formulas) rather than broken (by being ignored) seems, at least to me, to *clearly* demonstrate the constraining power of the constitution. Amendment formulas are themselves *part of the constitutional constraint*, and going through the motions of the amendment process demonstrates, quite clearly, that while a change to the constitution is needed, the constraining power of the constitution nevertheless remains intact. Once Ulysses is tied to the mast, it's not as though he is meant to be left there for the rest of his life: there is, presumably, some agreed upon (or, perhaps, intuitive

condition[206]) under which he is to be released. This condition of release is as much a part of the precommitment as is the constraint.[207]

The *breaking* of a constitution is normatively different from the undoing of a constitution, just as seeking release from a promise is normatively different from simply breaking it. A state which claims to be bound by a constitution, yet whose behaviour suggests a lack of such constraint, is a state which is (or ought to be) eyed with cautious reluctance. There is risk in striking military treaties or trade agreements with a state and expecting obedience or forbearance, particularly if that state seems incapable of (or disinterested in) meeting the conditions it sets for itself.[208] This seems to be the case regardless of whether a constitution is democratic or not. We don't *need* constitutions to be democratic to notice when an administration is not going along with them. This metric is as useful within the state as it is outside the state; it'd be irresponsible for foreign nations to make agreements with a state which demonstrates little recognition of its own constitutional affairs, let alone those that come from without. It's one thing for a state to be overtly and openly cruel and abusive of its people for gain in the international arena; it's another for them to feign care and principled concern while actively disregarding the very rules that purport to govern them.

---

[206] Here, I mean to include the possibly that, given the known desire to avoid jumping overboard when he hears the sirens sing, it can be reasonably determined by the crew that Ulysses may be untied once he can no longer hear the sirens – say, once they have sailed a certain distance away from their island, or once the ship is returned to the harbour – even if it hasn't been discussed beforehand.

[207] Certainly, it's possible for amendment to become a problem for constitutionalism, but not because of it happening; rather, it seems to me more likely to impair the constraining power of the constitution if, after repeated failure to pass a proposed amendment, people come to ignore the process because it's simply too difficult to successfully make the change. But it is not the existence of the amendment process that weakens the constraining power; rather it's the impossibility of constitutional amendment that will weaken constraining power.

[208] This isn't limited to the claim that 'the people' have chosen. Somebody makes choices for the state: whether it's the people or the government actors, the people who 'choose' what will be done are the people whose behaviour is to be assessed.

Democratic Legitimacy and Constitutional Precommitment

With all this, why the emphasis on claiming that constitutions represent the will of the people? I have a few thoughts on this.

First, there is the need to answer the binding-others objection and rectify the perceived tension in constitutional democracies. Since it is an assumption that commitments are only binding if they are chosen (in some manner) for oneself, it makes sense that we would ask whether constitutional commitments have been chosen for ourselves. The lacuna of commitment theory has resulted in a position where our only tools for dealing with this question are present in the language of democratic legitimacy: if all we have is a hammer, then we simply expand our conceptual category of what counts as a nail. Thus, the advocates argue for conditions under which the people can be said to have somehow chosen the constraints. Nevertheless, if commitment does not require an exercise of agency, then democratic legitimacy simply doesn't matter for constitutional precommitment.

But perhaps there's some other benefit to the broader constitutional game when constitutions aim to represent the will of the people. It seems true that a constitution that accurately represents the will of the people it governs should be more successful in governing; people who accept and see their aims and values reflected in their constitution are probably unlikely to revolt or leave. From the state's internal point of view, this will be *prima facie* good for stability and longevity: an enduring society will be closer at hand. Such stability may also offer benefit to the general, international level; no one wants to make treaties with a country constantly beset by civil war and revolution, since there is a significant risk of those agreements falling to the wayside. Avoiding revolt is desirable for the constitutional game, since the absence of revolt indicates at least some level of stability and reliability. It may be thought, as noted

above, that democracy is the best system of avoiding revolt, since presumably if the people are permitted to do what they want, then there isn't much 'rational' reason for them to complain.

At the same time, from the international perspective, constant constitutional change has the same effects to the international community whether broken or undone. And, insofar as different people have different minds, constitutions may require frequent amendment if they are to really and truly represent the will of the people. Frequently undoing one's constitutional commitments through formal amendment may be as troublesome as simply breaking one's commitments, even if we might tend to think of one as being more morally acceptable or 'legitimate' than the other. The parallel example here is the case of the person who makes frequent promises, but always needs to be released from them. A person who makes true promises but who frequently needs to be released from them (even through no fault of their own) is still an unreliable promiser, as is the person who makes promises they never intend to keep. We may certainly think the latter has some sort of moral failing that the former does not, but we'd be unwise to rely on either of them for anything of great importance, as we do not know from either whether our needs will be met.

A state with a constitution that is under constant revision is potentially as unreliable as a state which simply ignores the constitution that purports to govern them. Still, there *is* a difference. To go through the amendment process, rather than ignore or strike from the page some undesirable segment, requires a very particular kind of forbearance and recognition. The *point* of constitutionalism is the process, not the substance. Substance can shape and determine the process, but the demonstrative value of having constitutional constraints is to see *whether* one will abide by them rather than to concretely ensure that one does: constitutions bind by providing the opportunity to demonstrate whether one will be bound. A state who does not abide

by its constitution will have compromised the rule of law, as those most important in its function do not feel obliged to abide by its dictates. This is not undone by constitutional amendment but is uniquely demonstrated in the very process of amendment.

It might also be the case that a constitution which represents the will of the people is one which can better act as a check on state power since, to the extent that people can know their will, a constitution which represents the will of the people is a constitution that the people can know. A state governed by some set of secret principles, known only by the elite, will be harder for regular individuals to check: we can't determine whether state actors are toeing the line if the line is invisible to us. Again, the constitution won't be able to enforce itself. Constitutional precommitment as a useful tool requires that those judging the actions of the state can understand and comprehend the expectations that constitutional law sets out, but this means that the lines must be visible. If the people are unable to know the line – whether because it is described in arcane legalese, or because it is hidden throughout decades and centuries of case law – then they are unable to judge whether a line has been crossed.

Conclusion

In abstract, the purpose of a constitution is to aid in promoting stability, protection, and constraint. But the *point* of something like a constitution is to centralise those things that, as a matter of history, are known and have been shown to be destabilising or disruptive to the continued existence of society. If we're resolving to survive, and we need society to do so, then we are committed to certain things to protect and serve the longevity of society.

In truth, it seems to me that constitutions *should* be quite stable as there are few essential commitments that will change over time. One of the reasons constitutions are written in such abstract language is purportedly to avoid the need for constant revision and to avoid missing important details. But abstract language does little to teach us the lessons that we are supposed to

learn: it does nothing for us to know that we have, for instance, the right to liberty if liberty is left entirely undefined. Society seems hardly stable if it is fundamentally divided on conceptual disagreement about what we think we have the liberty to do: regardless of whether we're wrong about what rights we have, unhappiness festers when we think our rights are infringed.

Instead, constitutional language should be highly precise, as in other areas of law. There should be little to no room for misunderstanding or disagreement about what the constitution says. It should not be left up to interpretation what our constitution requires[209] any more than it should be left to interpretation what I have promised you. The need for frequent constitutional change should dissipate if it is indeed true that the commitments of stable society are predominately unchanging; frequent 'updating' should not be needed, as our commitments will be quite unchanging.

The end of survival commits us to certain things (and against certain others) which are necessarily done or avoided. These are not commitments we have because we think they are good or worthwhile; they are entirely conditional on survival and continued existence being an end we hope to achieve. A good constitution should center on those commitments; it should not be about forging some kind of national personality, or signaling our national mores.[210] It should modestly aim to keep track of the areas, globally and contingently, which have threatened the stability of the state. It should serve as a repository for important lessons from history that we would be woeful to forget.

---

[209] Importantly, "what the constitution requires" is not the same as "what is required to align with the constitution". We may disagree about what is the best protection against infringing on some group's freedom of religion, or freedom of association. We should not, however, be unclear about what freedom of religion or freedom of association are.

[210] This is not to say that a national identity cannot be created through the contents of a constitution. Rather, I mean only to say that the point of such an identity should be second to the resolution of stability and survival. While national identity may support that pursuit, creating that national identity should not be the end in itself.

We don't *need* constitutions for this end, ultimately; historical lessons could easily be passed on through storytelling, or mythology, or other. But it may be that these other formats complicate the success of passing along the lessons: storytelling may too easily confuse fact and fiction, and mythology may be too fanciful and imaginative. We don't need to entrench them in a constitution or recognise them as law, but chances are high that history will repeat itself if they are not kept in public consciousness somehow. While constitutions are not the only way to centralise these commitments, there might certainly be reason to think it's (so far) the best.

Bibliography

Allan, James "Thin Beats Fat Once Again – Conceptions of Democracy" *Law and Philosophy 25* (2006): 533-559.

Aquinas, St. Thomas. Summa Theologica. Translated by Fathers of the English Dominican Province. California: Coyote Canyon Press, 2018.

Bellamy, Richard, and Dario Castiglione. "Constitutionalism and Democracy - Political Theory and the American Constitution." *British Journal of Political Science* 27, no. 4 (1997): 595–618. http://www.jstor.org/stable/194230.

Comella, Victor Ferreres. Constitutional Courts and Democratic Values: A European Perspective. Yale University Press, 2009.

Consani, Cristina Foroni. "Pré-Compromisso Constitucional e Autonomia Coletiva: Uma Conciliação Possível?" *Revista de Estudos Constitucionais, Hermenêutica e Teoria Do Direito,* 7, no. 3, (2015): 235–242. https://doi.org/10.4013/rechtd.2015.73.03.

Dippel, Horst. "Modern Constitutionalism, an Introduction to a History in Need of Writing," Tijdschrift voor Rechtsgeschiedenis/Legal History Review 73, no. 1 (2005): 153-170.

Dodek, Adam. *The Canadian Constitution*. Dundurn Press. 2016.

Dworkin, Ronald. *Law's Empire*. Harvard University Press. 1986.

Dworkin, Ronald. "The Moral Reading and the Majoritarian Premise" In *Deliberative Democracy and Human Rights* edited by Ronald C. Slye, 81-115. Yale University Press, 2008. https://doi.org/10.12987/9780300128734-007

Elster, Jon. Ulysses and the Sirens: Studies in Rationality and Irrationality. Cambridge University Press. 1979.

Elster, Jon. Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints. Cambridge University Press. 2000.

Freeman, Samuel. "Constitutional Democracy and the Legitimacy of Judicial Review." *Law and Philosophy* 9, no. 4 (1990): 327–70. http://www.jstor.org/stable/3504771.

Habermas, Jürgen, and William Rehg. "Constitutional Democracy: A Paradoxical Union of Contradictory Principles?" *Political Theory* 29, no. 6 (2001): 766–81. http://www.jstor.org/stable/3072601.

Hamilton, Alexander, James Madison, and John Jay. *The Federalist Papers: A Collection of Essays*. Coventry House Publishing, 2015.

Hart, H. L. A. "Positivism and the Separation of Law and Morals." *Harvard Law Review* 71, no. 4 (1958): 593–629. https://doi.org/10.2307/1338225.

Hart, H. L. A. *The Concept of Law*. Second Edition. Clarendon Press. 1997.

Haslanger, Sally "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" *Noûs* 34, no. 1 (2000): 31-55. https://www.mit.edu/~shaslang/papers/WIGRnous.pdf

Hayek, F. A. The Collected Works of F. A. Hayek. Vol. XVII, The Constitution of Liberty: The Definitive Edition. Edited by Ronald Hamowy. Routledge. 2011.

Hobbes, Thomas. Leviathan. Edited by Christopher Brooke. Penguin Books Ltd, 2017.

Holmes, Stephen "Precommitment and the Paradox of Democracy." In *Constitutionalism and Democracy*, edited by Jon Elster and Rune Slagstad (Cambridge: Cambridge University Press, 1988), 227.

Hutchinson, Allan C. *Democracy and Constitutions: Putting Citizens First*. University of Toronto Press. 2021.

Jefferson, Thomas. "From Thomas Jefferson to James Madison, 28 August 1789", Founders Online, National Archives. https://founders.archives.gov/documents/Jefferson/01-15-02-0354.

Johnston, David. Foreword to *The Canadian Constitution*, Edited by Adam Dodek, 9-10. Toronto: Dundurn Press. 2016.

McLachlin, Beverley. Introduction to *The Canadian Constitution*, Edited by Adam Dodek, 11-12. Toronto: Dundurn Press. 2016.

Nippel, Wilfried. Ancient and Modern Democracy: Two Concepts of Liberty? Cambridge University Press. 2016.

Orwell, George. *Animal Farm,* Penguin Books. 1951.

Rothwell, Christina. "The Constitution of Theseus: An Investigation into the Metaphysics of Constitutional Precommitment." M.A. Thesis, McMaster University, 2017. http://hdl.handle.net/11375/22276.

Shpall, Sam "Wide and Narrow Scope." *Philosophical Studies* 163, No. 3 (2013), 717-736. https://www.jstor.org/stable/41932698

Shpall, Sam. "Moral and Rational Commitment." *Philosophy and Phenomenological Research* 88, no. 1 (2014): 146–72. http://www.jstor.org/stable/24673002.

Vile, M. J. C. *Constitutionalism and the Separation of Powers*. Second Edition. Liberty Fund, Inc. 1998.

Waldron, Jeremy. *Law and Disagreement*. Oxford University Press. 1999.

Waluchow, W. J. "Constitutions as Living Trees: An Idiot Defends." *Canadian Journal of Law & Jurisprudence* 18, no. 2 (2005): 207–47. doi:10.1017/S0841820900003994.

Waluchow, W. J. *A Common Law Theory of Judicial Review*. Cambridge University Press. 2007.