Theorizing Justice for a Realistic Utopia: The Methodological Implications

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities

2015

Stephanie Rinaldi

School of Social Sciences
Contents

Introduction........................................................................................................................................... 8

Theorizing Justice for a Realistic Utopia: An Overview................................................................. 8
  0.1 Essentials of the RU Methodology ......................................................................................... 13

The Structure of a Theory of Justice ............................................................................................. 16

Chapter Summaries .......................................................................................................................... 19

1. Realistic Utopianism ................................................................................................................. 25
  1.1 Dividing Methodologies by Their Aim .................................................................................. 27
    1.1.1 Non-Ideal (NI) Approaches ......................................................................................... 34
    1.1.2 Realistic Utopian (RU) Approaches ........................................................................... 36
    1.1.3 Utopian (U) Approaches ............................................................................................ 39
    1.1.4 Theory of Ideals (ToI) Approaches ............................................................................. 41
  1.2 The Realistic Utopian Approach ............................................................................................ 43
    1.2.1 RU Idealisations ........................................................................................................ 44
    1.2.2 Feasibility ..................................................................................................................... 45
    1.2.2.1 An Uncontroversial Conception of Feasibility ....................................................... 49
    1.2.3 Value ............................................................................................................................ 53
  1.3 Existing Realistic Utopians .................................................................................................... 55
    1.3.1 The ‘Core Commitment’ and Existing RU Theorists .................................................. 56
    1.3.2 Interpretations of Feasibility Constraints ..................................................................... 58
      1.3.2.1 Contextual Factors ............................................................................................... 59
      1.3.2.2 Psychological Factors .......................................................................................... 62
      1.3.2.3 Institutional Factors ............................................................................................ 65

2. Contextualism .............................................................................................................................. 70
  2.1 Contextualism ......................................................................................................................... 71
  2.2 Contextualism Over Universalism ......................................................................................... 76
    2.2.1 RU Theory and Determinacy ...................................................................................... 78
    2.2.2 Determine Principles and Contextual Facts ............................................................... 81

2.3 Argument in Favour of Weak Contextualism ........................................................................ 85
Tables and Figures

Figure 1: The Structure of a Theory of Justice ........................................16

Figure 2: Functions and Aims of Conceptions of Justice ......................30

Table 1: Forms of Contextualism.........................................................85
Abstract

Theorizing Justice for a Realistic Utopia: The Methodological Implications

Many political theorists agree that practical recommendations are a valuable attribute of a theory of justice. For some theorists, who I refer to as Realistic Utopians, this means that they should aim to develop principles of justice to govern the constituents of a moderately idealised society. In this thesis I establish the methodological implications of the Realistic Utopian approach and identify their unifying theme.

The Realistic Utopian approach to developing principles of justice has received widespread support in the literature. Considering an idealised society allows theorists to pursue aspirational recommendations of justice, whilst the moderate nature of the idealisations ensures that those recommendations are in some sense practical. Their aim represents the idea that it is important to develop principles that describe how just we can possibly be. The popularity of this aim has resulted in many theorists adopting a broadly similar approach. It permits theorists to use idealisations in their deliberations, but requires these idealisations to be limited by feasibility constraints. In this thesis I explain the core commitment shared by all RU theorists, distinguish the RU approach to justice from alternative approaches, and then defend three additional methodological requirements that RU theorists must adopt. Each of these three additional methodological requirements is entailed by the shared core commitment to develop principles to govern a moderately idealised society, and each imposes more stringent requirements on the methodology than proponents of the view typically acknowledge. I do not claim that this list is exhaustive—there may be further methodological requirements that particular RU theorists choose to adopt in addition to those set out here. My claim is that the three requirements I identify impose restrictions on suitable methodologies for developing principles of justice that must be respected by any theorist who wants these principles to apply to a moderately idealised utopia.

The idea that the grounds of justice should be sensitive to the views of those to whom the principles may eventually apply unifies these requirements. Although Realistic Utopian principles of justice are not designed to apply immediately, the methodology requires they be designed with a particular context in mind. Meeting this requirement via sensitivity to the views of those within the context allows theorists to meet all three methodological requirements simultaneously.

Stephanie Rinaldi

The University of Manchester

Submitted for the degree of Doctor of Philosophy

25th February 2015
Declaration

The author declares that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
Introduction

Theorizing Justice for a Realistic Utopia: An Overview

Political theorists interested in justice ask what morality requires of individuals living together in a cooperative society. What’s fair, what we owe to each other, and how we should treat one another are all questions that motivate theorists interested in the concept. Many think that asking these kinds of moral questions is valuable because doing so can help theorists to develop practical recommendations for how we ought to live our lives. The view that ‘action-guiding’ recommendations are important is widespread. Onora O’Neill says that theorists should pursue ‘reasoning which we and others can use in personal and public life not merely to judge and appraise what is going on, not merely to assess what has been done, but to guide activity’ (O’Neill, 1996, p. 3). Scanlon, Nagel, and others describe political theorists’ aim similarly: ‘the point of judgements about right and wrong is not to make claims about what the spatiotemporal world is like. The point of such judgements is, rather, a practical one: they make claims about what we have reason to do’ (Scanlon, 1998, p. 2). ‘One should think of political theory as an enterprise of discovery—the discovery of human possibilities whose coming to actuality is encouraged and assisted by the discovery itself. That is certainly how most of the traditional figures of political theory have seen it’ (Nagel, 1991, p. 6). Although these action-guiding recommendations are the goal of many political
theorists, the sense in which they guide action is slightly different, depending on what the theorist is trying to achieve. For some, theories are bounded by the constraints of the real-world in order for their recommendations to apply to us. Others, those who adopt a different interpretation of the kind of political theory that makes these recommendations—in which principles are developed for citizens of a hypothetical but practically possible world—I call Realistic Utopian (RU). RU theorists take theorising at this level of abstraction to be (at least part of) what political theorists should be interested in.

The central claim of the thesis is that the RU approach is subject to more stringent methodological commitments than existing proponents of the view typically acknowledge. I argue that their approach should be constrained in various ways in order to yield the kind of recommendations they’re interested in. These constraints are unified by the idea that RU theories of justice must be sensitive to what the people to whom the principles of justice ultimately apply think.

Rawls coined the term ‘realistic utopia’ and describes thinking about justice in this way as asking ‘what a perfectly just, or nearly just, constitutional regime might be like, and whether it may come about…under realistic, though reasonably favourable, conditions’ (Rawls, 2001, p. 13). In Law of Peoples, where he develops and uses the phrase most precisely, Rawls is referring to a particular, substantive utopia of ‘a reasonably just constitutional democracy’ (Rawls, 1999, p. 17). I, on the other hand, use the phrase in an adjectival sense, to refer to a particular kind of utopian theorising. As such, my understanding and use of the phrase is intended to be broader than Rawls’s in Law of Peoples. According to my taxonomy, the defining characteristic of RU theorists is a commitment to understanding principles of justice to provide action-guiding recommendations that given favourable but possible circumstances we should be able to achieve. I call this the ‘core RU commitment’. The core RU commitment entails that RU theorists are committed to both realism and utopianism in some sense. The principles they develop govern the constituents of a moderately idealised society.
The core RU commitment and the broad methodology it entails are shared by many existing theorists. We can take these areas of agreement to establish a working definition of the approach expressed as a claim about what a theory of justice should strive to do. The working definition I start with says that ‘RU theorists design action-guiding principles for the constituents of a moderately idealised society’. Throughout the thesis I will develop the methodological requirements to meet this aim and by the conclusion have the richer, fully unpacked, and spelled-out definition of the precise meaning. The methodology requires principles to be ultimately implementable in order for theorists to plausibly say that they are action-guiding. Many theorists interpret feasibility constraints to be a necessary condition of meeting this requirement. Conceptions of feasibility are disparate but converge on the following ideas: principles should be compatible with the true laws of science; the circumstances of justice; and facts that make their successful completion extremely unlikely conditional upon the relevant agent(s) trying. I discuss this in further detail, but for now it will suffice to recognise the core RU commitment as capturing the intuition that we should be thinking about justice in the world as we roughly know it, and for people as we roughly know them. The ‘roughly’ qualifier in this description draws attention to the fact that the purpose of RU conceptions of justice is to specify an ideal. This allows their deliberations to be free from the worst features of the world and humans living amongst one another as we currently know them; theorists are permitted to make some idealisations during their deliberations.

While the core RU commitment is shared by many theorists, there is much disagreement among these theorists over the more specific methodological commitments that flow from this. The aim of this thesis is to show that the core RU commitment entails a much narrower set of more specific methodological commitments than is commonly thought to be the case. I establish these commitments to develop a fuller definition of the RU approach to justice but stop short of developing a full methodology. I refer to RU

---

1 I list numerous examples of these in the first chapter, paying particular attention to Rawls’s (1999) Original Position and Dworkin’s (2000) Hypothetical Insurance Market.
theorists’ endeavour as the RU ‘approach’ to indicate that my concern is with different theorist’s strategies to develop a conception of justice understood according to the core RU commitment, and not to make claims about the content of those conceptions themselves. In particular, I claim that theorists should make RU conceptions of justice sensitive to i) their context, ii) their stability, and iii) their legitimacy as well as the broad commitment to feasibility that existing proponents already share. These commitments are ‘more determinate’ in the sense that they either demand particular methodological or justificatory principles be included in the methodology or impose constraints so that fewer such principles are compatible with the approach. Indeterminacy regarding these matters is incompatible with the core RU commitment and so RU theorists must take these things into account. They must accept that each of these commitments is a necessary condition of producing an ultimately implementable conception of justice despite the fact that many existing theorists who share the core commitment either do not recognise them or reject them as necessary conditions of justification. In addition, I specify the way in which theorists must take these things into account, since some interpretations of these requirements ought to be rejected on methodological or substantive grounds. Existing RU methodologies are under-theorised when theorists have mistakenly failed to recognise these concepts as fundamental to their aim. In other cases, existing RU methodologies conflict with my claims. Some RU theorists have made what I take to be false claims regarding the significance and/or proper interpretation of contextualism, stability, and legitimacy within their theories. These claims are addressed as I come across them. The arguments I make in favour of each of these methodological commitments draw upon the function and aim of justice that defines RU theorising—to provide action-guiding recommendations for a realistic utopia. They are committed to producing a conception of justice that can guide action measured by whether or not it is ultimately implementable.

These commitments may be unified by attentiveness to people’s views about justice. Theorists may fulfil each methodological requirement in a number of different ways. For example, the first requirement (of contextualism) can be met by sensitivity to a number of different kinds of contextual fact, including
people’s views. As my discussion of the additional requirements progresses we will see that it may be the case that people’s views about justice are the preferred type of contextual fact because sensitivity to them permits theorists to meet each requirement simultaneously. This idea—that sensitivity to the views of those to whom principles of justice may eventually apply—is considered more tentatively, and I don’t take it to be my central contribution.

Considering the details of the RU approach and developing a more fully worked-out methodology is important because developing principles to govern a moderately idealised constituency is an important task for political theorists. To develop such principles is to explore how just we can possibly be within the limits of practical possibility. The results of this approach are valuable to inform what justice requires of us here and now, and to see what we can do to change the world for the better in this respect. Rawls describes action-guiding recommendations as providing ‘a long-term goal of political endeavour’ and that ‘in working toward it [we can] give meaning to what we can do today’ (Rawls, 1999, p. 128). This idea is the most plausible understanding of the instrumental value of RU theorising. It says that developing an RU conception of justice is valuable because doing so is instrumentally valuable to making the world a better, and more just, place. Indeed, on this view, ‘our hope for the future of society rests on the belief that the social world allows at least a decent political order’ (Rawls, 2001, p. 4).

For other theorists, the RU approach is intrinsically valuable because its results reflect a conceptual truth about the nature of justice. One view is that developing guidance for the constituents of an idealised rather than actual society is necessary to give the theorist a critical insight into the concept of justice that she would struggle to achieve if she were to consider non-ideal circumstances only. In this case, the meaning of justice can only be found if we abstract away from some the circumstances that we are faced with, but recognise that others will always constrain us. Regardless of the precise reasons many existing theorists adopt the approach, the methodological investigation of this thesis is valuable in virtue of the fact that many theorists do endorse it. The claims I make in this thesis provide a basis from which existing RU
theories can be critiqued and compared by those who are not themselves moved to develop principles for a moderately idealised society. Given this value of the approach, sensitivity to considerations of feasibility, context, stability, and legitimacy is important because ignoring them is inconsistent with the aim of the RU theory to develop principles to guide a moderately idealised constituency. In this way, the thesis is an investigation into the approach itself.

The RU approach’s being valuable does not preclude the idea that developing requirements beyond the frontier of practical possibility or within the constraints that we face here and now are not also important. There are, of course, alternatives to the particular RU aim when it comes to developing a theory of justice. Alternative theorists fall into one of two categories. They either view the function of principles of justice to be to reflect the truth of the ideal, or they agree that theorists should produce action-guiding recommendations, but disagree over the constituency principles should apply to in order to fulfil their particular aims. Alternatives to the RU approach are not necessarily incompatible and we might consider whether a combination of approaches is ultimately preferable (Carens, 1996, p. 156).

0.1 Essentials of the RU Methodology

RU theorists produce recommendations for the constituents of a moderately idealised society. In this sense their approach differs from those who take the meaning of ‘action-guiding’ to imply recommendations we can comply with here and now (I call these theorists Non Idealists). The use of idealisation in the RU approach reflects this difference. An idealisation is a falsehood about the empirical nature of the situation to which theorists are prescribing principles. Rawls, for example, prescribes principles for citizens inhabiting a ‘well-ordered’ society. Such a society is effectively regulated by a public conception of justice. This means that both its institutions and citizens are regulated and motivated by the conception of justice that orders them (Rawls, 2001, pp. 8-9). This

---

2 Some theorists reject the use of idealisations and instead only permit abstractions. Here, the two words should be understood interchangeably; a theorist’s use of abstraction can reveal her commitment to the RU approach in the same way that her use of idealisations can. I outline the reason that I reject the distinction between idealisations and abstractions in the first chapter.
situation clearly differs from the matters of fact in the real world with regards to at least some individuals. RU theorists may make such idealisations and then deliberate over the appropriate principles of justice for the idealised society. Idealising the nature of the empirical world so that the constituency of their recommendations is utopian is necessary to the RU theorists’ aim.

This definitional feature of the approach is adopted and shared by all existing proponents. Theorists may disagree on the reasoning in favour of idealisations, but they share the idea that they are in some sense indispensable. They also, as RU theorists, share the view that there must be a limit upon which facts can plausibly be idealised for the results of the approach to still fulfil the stated aim, although they disagree over where this limit ought to be set. The literature is lacking in terms of a general characterisation of a consensus over this limit. I argue in the first chapter that the most plausible interpretation of RU theorists’ aim means that they ought to subscribe to the following rule: idealisations are permissible so long as they do not undermine a good-faith interpretation of the action-guiding nature of their principles. An idealisation undermines the action-guiding nature of principles when their guidance could never be applicable in the real world. For example, idealising humans so that they couldn’t feel pain would yield principles that could never be applicable in the world as we know it. Recall that the aim is to produce principles that push the limits of the practically possible. In order to meet this aim, RU theorists must consider principles that could ultimately be implemented into the world; this represents a good-faith interpretation of what makes a principle action-guiding. This also matches the way most theorists limit their idealisations, although it isn’t always stated as such in their methodology.

Limiting idealisations to ensure that principles are sufficiently action-guiding is permissive and requires further analysis and argument to determine whether particular examples of idealisations would be permitted or not. The fact that the general criterion permits a large number of idealisations is not to say that a largely idealised utopia is the approach that the majority in the literature favour. Existing RU theorists provide the rationales for particular idealisations to show that they don’t violate the limitation. The limitation also does not specify any
particular idealisations to be necessary to produce a sound RU theory. Decisions about particular idealisations may be made by particular theorists and will depend on what they take to be useful to their end of providing action-guiding recommendations.

This idea of a limitation to idealisations is an important requirement of the approach and distinguishes RU theorists from their more utopian alternative (I call these Utopian theorists). Existing RU theorists have converged over the idea that principles must be feasible in order to meet this requirement. Utopian theorists, on the other hand, reject the idea that requirements of justice can be limited by empirical concerns in this way. The concept of feasibility is appealing for RU theorists since it reflects the intuitive ideal that they’re developing principles for the limits of the practically possible. As such, existing RU theorists take feasibility to be a necessary feature of a conception of justice being ultimately implementable and acknowledge that this imposes a limit to their use of idealisations.

Feasibility constraints specify the sense in which the society to which RU recommendations apply is realistic. Theorists recognise that certain empirical circumstances block normative requirements. For example, circumstances that make an action impossible block the moral requirement to perform that action. If it is impossible for the state to make sure that each one of its citizens is as happy as she could possibly be, then morality cannot require the state to do that. On the other hand, empirical circumstances might block actions if they render the completion of those actions infeasible. It would be possible for the state to give each and every citizen her own Beagle puppy (to make her happy), but it’s clearly not a feasible request. I take feasibility to impose more restrictions than possibility (so that the realm of feasibility is included within, but does not constitute the limits of, the realm of possibility). If certain empirical circumstances are persistent and universally block requirements in this way in the world as we know it, then theorists should not idealise them. As such, feasible implementation is a necessary condition of a sound RU theory of justice and infeasible principles are not plausible candidates. The working
definition of the RU approach can incorporate this so that it is clear that infeasible principles of justice cannot be RU principles of justice.

The Structure of a Theory of Justice

Here, I outline my interpretation of the structure of a political theory methodology and how it fits into a wider theory of justice. Doing so allows me to more precisely articulate the claims that I make in forthcoming summaries and chapters. On my understanding, a theory of justice is composed of three types of principles: methodological, justificatory, and substantive. The focus of the thesis is, for the most part, the first two types of principles (which together constitute the methodology) and so it is instructive for me to explain their natures and relationship. A justificatory principle might, for example, stipulate that ‘sound theories of justice must be feasibly implementable’. A methodological principle, on the other hand, would describe how to meet the requirement by saying ‘follow this procedure to yield feasible results’. As I discuss shortly and elaborate in chapter one, existing RU theorists share some justificatory principles. The claims I make in the rest of the thesis require any plausible version of the RU approach to contain particular justificatory and methodological principles.

The methodology is made up of methodological and justificatory principles. The conception of justice, i.e. the guidance on what justice requires us to do, is made up of substantive principles and all three together constitute a full theory. This relation is diagrammatically represented below.

![Diagram of the Structure of a Theory of Justice]

*Figure 1. The Structure of a Theory of Justice*
Conventionally, theorists understand the modifier ‘methodological’ to describe principles that ‘specify the generation of substantive principles’ (Cohen, 2003, p. 221). So, rather than describe actions one is required to perform as a matter of justice, methodological principles describe the way in which theorists settle on the content of those actions. I don’t disagree with this view simply stated, but we need to say more on what, exactly, it means for a principle to tell us how to choose other principles.

Significantly, it’s unclear whether this understanding implies that methodological principles are normative. Cohen describes the principle embodied in Rawls’s original position (1999) as methodological (ibid.). The principle he has in mind here presumably says something like ‘theorists should adopt [substantive] principles that would be selected by parties in the original position’. The ‘should’ in this description might lead us to conclude that such a principle is normative, since it issues a prescription. However, for Cohen, the very purpose of labelling the principle methodological is to distinguish it from what he calls normative principles. This suggests that, in fact, methodological principles lack normative content. This apparent ambiguity is settled by distinguishing between ought claims regarding what justice requires we do and ought claims about what theorists should do in order to develop principles of justice. The former is what makes a principle normative on Cohen’s view.

On my view, although it’s plausible that methodological principles could include the second kind of normative content, it is clearer to separate normative and non-normative principles within political theory methodology since there are plausible roles for both types of principle. Methodological principles are non-normative when they tell theorists how to satisfy or generate substantive or justificatory principles without including any ought claims about whether they should. When I use the term ‘methodological principle’, I will have this non-normative meaning in mind. An example of a non-normative methodological principle is the one that describes the conditions of the original position. It gives details on the technique that theorists are to employ to satisfy the procedure (when the procedure itself is ‘a device that maps a fair and impartial point of view’). This kind of principle is not ‘free-standing’. That is, it
has to be backed up by a justification (Ronzoni & Valentini, 2008, p. 409). In this example such a justification would explain why Rawlsian conditions are the most appropriate for the original position choice procedure and would themselves be an example of a justificatory principle.

Principles that issue normative prescriptions on how to develop a political theory I call justificatory principles. They specify the requirements of a sound theory of justice. They tell us which moral and non-moral facts ground substantive and methodological principles of justice or, in some cases, they specify a procedure with right-making properties. Since they inform what makes a substantive principle normatively sound, they themselves are normative in the second sense described above. The principle about the Original Position that Cohen described as methodological—‘theorists should adopt principles that would be selected by parties in the original position’—is, according to my terminology, a justificatory principle.

Together, methodological and justificatory principles supplement one another and constitute a methodology. Justificatory principles specify what theorists ought to do, and methodological principles refine the actual approach that they take to ensure that the justificatory requirements are met (as we saw with the Original Position example).

Clearly, a methodology isn’t all that’s required in a theory of justice. As I’ve mentioned, adopting a method yields substantive principles (this is represented by the line in figure 1 that connects justificatory and methodological principles to substantive ones). Substantive principles make up the conception of justice. The principle ‘each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all’ (Rawls, 2001, p. 42) is an example of a substantive principle. In other words, a conception of justice is a set of action-guiding, normative, principles conformity to which achieves justice (this conforms to Cohen’s use of the term normative principles; the first sense of normative that I described above). Aside from broad assumptions in favour of some substantive values (freedom, equality, respect, etc.), I am not concerned with substantive claims of justice in this thesis.
Chapter Summaries

Chapter one is devoted to the definition, function, and aims of the RU approach to developing principles of justice. I establish a working definition of the approach by considering the best way to categorise political theory methodologies. I distinguish two fundamental functions of justice: to produce a set of action-guiding recommendations (this function can be interpreted in various ways), or to uncover the ideal, or truth, of the concept of justice. RU theorists adopt the former, and have a particular interpretation of the aim to develop action-guiding recommendations, as we have seen. This interpretation distinguishes them from alternative approaches to developing action-guiding recommendations and informs the essentials of their methodology. Their use of idealisations distinguishes them from the Non-Ideal interpretation of action-guiding recommendations, and the constraint to develop a conception of justice that can be ultimately implemented distinguishes them from the Utopian alternative at the other end of the spectrum. I demonstrate RU theory’s initial plausibility and prevalence in the literature, and show why RU theorists’ shared commitment to the function of justice entails the broad methodological commitments outlined above.

The methodological commitment to feasibility raises a number of questions regarding what, exactly, it means to say that a particular action is feasible. For example, it is unclear whether a current, but alterable, non-moral fact counts as a feasibility constraint. Does the fact that, given existing distributions, we lack the spare resources to supply everyone with tertiary education, mean that doing so isn’t a requirement of justice? Further difficulties arise when we try to specify what makes a fact ‘alterable’ in the relevant sense. Is the possibility of redistribution sufficient to say that the existing distribution is alterable? What if those with ‘spare’ resources are using them for some other morally worthy cause? What if there is no fair way of redistributing now, but there might be in the future? Should we regard the future as the next five years or the next five-hundred when making these assessments? Theorists disagree over these kinds of questions and thus endorse competing conceptions of feasibility. One example of such a disagreement in the literature is over the question of
whether moral costs should be included in feasibility assessments. That is, whether the moral damage imposed merely by changing social ideals (caused by forcing agreement or having to change people’s minds, for instance) should be factored into the feasibility assessment. Gilabert and Lawford-Smith discuss theirs and opponents’ views on this in *Political Feasibility: A Conceptual Exploration* (2012, pp. 816-7). The extent of disagreement amongst RU theorists on issues like these means that the only aspects of a conception of feasibility they share are those that rules out actions that are incompatible with either i) the true laws of science, ii) the circumstances of justice or iii) a reasonable chance of success conditional upon trying. Although this is a somewhat thin conception of feasibility, it does mean that RU principles can’t require us to do things like turn back time (violates i), become totally altruistic (violates ii), or successfully dam a river where there is very little chance of us being able to do so (violates iii). It does appear to leave open a wide range of possibilities, for example to redistribute tertiary education (as I discussed above). Of course, there is room for dispute when it comes to interpreting even these uncontroversial conditions; inevitably there are borderline cases that are difficult to classify. This conception does, however, capture a number of our most firmly held intuitions regarding what it means for an action to be feasible. In addition to its intuitive appeal, sensitivity to these kinds of facts matches the way in which feasibility constraints are commonly used in the literature. As such, we can further refine the working definition of the RU approach to reflect this end. At the end of chapter one I discuss the concept of feasibility in some detail and several possibilities that we might consider if we were to attribute a thicker conception to RU theorists.

The broad methodological commitments (to feasibility and idealisation) of the RU approach are broad in the sense that they do not require RU theorists to adopt any particular idealising assumptions within the constraints of the feasible. As such, they don’t describe a particular RU methodology, but rather establish the boundaries of the approach. Despite these constraints and assumptions that many existing RU theories share, there are many more over
which theorists disagree.\(^3\) These disagreements lead to difficulties when it comes to characterising the approach in general terms; the shared basis of RU theory—commitment to a particular function and aim of conceptions of justice—does not entail a determinate methodology. That is to say, it does not prescribe to theorists a precise way of developing and refining their substantive principles of justice. It is not clear, for example, whether an RU theorist should take current facts about scarcity to impose a feasibility constraint, or whether she is permitted to idealise them away. It is to this indeterminacy that I speak in the thesis. Despite the absence of consensus over such a methodology, we can still identify one that does the best job. Consensus over the core RU commitment and the bare essentials of the type of approach that can direct us to this aim makes RU theory distinctive and gives the label meaning. There are, however, too many distinct methodologies to consider and not enough space or established parameters on the basis of which to judge the most successful existing RU approach. Instead, I establish what I take to be the minimum requirements of any successful RU approach. These are more stringent than the existing area of methodological consensus. The preceding discussion over the differences regarding appropriate idealisations and feasibility constraints invites investigation into these narrower constraints and sets the stage for the remaining three chapters.

It is the RU function and aim of a theory of justice that gives theorists reason to acknowledge these additional methodological commitments in the same way they do feasibility constraints. I agree with the necessity of the essentials of the method established by consensus in the first chapter, but think that more commitments stem from the core RU commitment. I take my task to be an

---

\(^3\) The way in which the literature delivers different results depending on adaptations to particular idealisations is highlighted by the following comment: ‘some tinkering with Rawls’ comments about the family gives you liberal feminism (à la Okin, 1989), some tinkering with Rawls’ assumption about society being closed gives you global justice (à la Beitz, 1979), some tinkering with the assumption that all people fall within the normal range of functioning gives you an account of just health care (à la Daniels, 1985) and some tinkering with Rawls’ assumption that society is culturally homogeneous gives you a liberal account of multiculturalism (à la Kymlicka, 1989)’ (Farrelly, 2007, p. 847). Each of these ideas emerges from disagreements with Rawls about the appropriate scope of idealisations or constraints. It demonstrates the idea that there is a range of possible approaches within these limits and that theorists may disagree which is the most plausible.
internal investigation into the commitments of realistic utopianism. My conclusion is that certain methodological requirements must be incorporated into any RU theory for the results of it to accurately fall under the definition of an RU conception of justice according to their view of the function and aim of a conception of justice. Once RU theorists acknowledge these requirements, there is still scope for variations within the approach (although the scope is narrowed).

The aim of the second chapter is to establish the first methodological constraint that is not widely adopted by existing proponents. In it, I say that the content of some substantive principles in an RU theory of justice ought to be partly justified by the context to which those principles are to apply. This claim requires RU theorists to adopt a contextualist justificatory principle into their methodology so that contextual facts form part of the grounds for some substantive principles. These contextual facts could reflect things such as scarcity, prevalent values, existing practices, people’s views, or social history. RU theorists should acknowledge the necessity of this justificatory requirement because sensitivity to contextual facts is necessary to develop ultimately implementable principles. Without such sensitivity, their principles fail to be sufficiently determinate. I also defend the particular type of contextualist principle that I propose—weak contextualism—against alternatives by showing that those alternatives are implausible.

In the third chapter, I demonstrate that RU conceptions of justice must be stable for the right reasons. If the conception cannot be shown to be stable in this way, then it is not an ultimately implementable conception of justice. A conception is stable for the right reasons when each citizen supports its recommendations so that it guides her actions without the threat of coercive force. Not only must this be possible, but theorists must show their conception of justice to be capable of guiding and supporting institutions and citizens over time. Theorists are committed to acknowledging the normative value of stability for the right reasons because it’s implausible to think that the feasibility of implementing a conception is normatively important but that its maintenance is not. If the requirement to produce ultimately implementable
principles commits theorists to acknowledging feasibility constraints, then it also commits them to acknowledging stability constraints. The second aim of the chapter is to show that the only plausible way that RU theorists can meet the justificatory requirement of stability is to make their conceptions of justice publicly justifiable. This means that a theorist must show that the content of her conception could be the subject of agreement of reasonable people motivated by their sense of justice and adopt a justificatory principle into her method as such. Alternative sources of stability are incompatible with the RU approach.

The final methodological claim, which I make in chapter four, says that RU conceptions of justice must be *legitimately* implementable. A necessary part of an RU conception’s justification is that it be legitimately implementable. This is because the nature of the core RU commitment requires theorists to respond to the fact that political power is permanent and coercive (this is true even when the societies that they are considering are stable for the right reasons). Such power is permanent because it’s necessary to order a cooperative society and it’s always coercive because citizens, even if they are motivated to comply, need assurances that others will do the same. Its coercive nature is problematic for the just implementation of a conception. Accounts of legitimacy justify coercion so that the implementation of a conception is not wrongful so long as it is legitimate. As such, RU theorists must defend and adopt a conception of legitimacy and show that their conception of justice and wider methodology are compatible with it. I don’t defend a particular account of legitimacy, but I do discuss some prevalent options and conclude that most are compatible with the RU approach.

At the end of the second (contextualism) chapter, I consider the differing sets of contextual facts that RU theorists might use to fulfil the requirement. I tentatively suggest that people’s views about justice are a plausible candidate. That what the people think about justice—where the people are defined as those to whom the principles are to apply—should be the contextual fact that forms a part of the grounds for those substantive principles. At the end of each of the following chapters I return to this suggestion and consider whether
this particular empirical sensitivity is compatible with each claim. As the thesis progresses, the suggestion develops strength and I conclude that this suggestion may unify each of the methodological requirements I’ve defended and thus that considering this kind of empirical data may be a good starting point for RU theorists when they begin their deliberations.
1. Realistic Utopianism

Introduction

In this thesis I argue that a commitment to the Realistic Utopian (RU) approach to justice has more determinate implications for the principles of an RU theory than proponents of the approach typically acknowledge. To say that the RU approach is more determinate in this way is to say that the range of methodological, justificatory, and substantive principles consistent with it is much narrower than is generally thought to be the case.

In this chapter I provide a working definition of the RU approach. The definition starts with the ‘core’ RU commitment to produce conceptions of justice that provide action-guiding recommendations for the constituents of a moderately idealised society. Developing a conception of justice for this kind of society commits RU theorists to various methodological constraints; it means that their conceptions of justice must be ‘ultimately implementable’. Ultimate implementability requires that it be possible for conceptions of justice to eventually govern the real world. Because they need only ‘eventually’ be able to do this, the world that theorists consider may be idealised to some extent—it could be radically different from the world as we currently know it. The limits of these idealisations are given by feasibility constraints. Existing RU theorists agree that feasibility informs the formulation of conceptions of justice and infeasible conceptions must be rejected. They also agree over an uncontroversial conception of feasibility that says that such conceptions must
be compatible with the laws of science, the circumstances of justice, and have a reasonable probability of success. These methodological commitments define the range of substantive, justificatory, and methodological principles that RU theorists typically take to be consistent with the approach; they establish constraints for such principles, rather than prescribing particular ones. I endorse them, but as we shall see in later chapters, think that they are insufficiently determinate to capture what’s required to produce the kind of ultimately implementable recommendations that RU theorists are interested in. As such, RU theorists should acknowledge three additional methodological constraints to contextualism, stability, and legitimacy. These refine the definition of the RU approach to be more precise and determinate. It still, however, permits a range of methodologies within the constraints. Anyone of these methodologies yields substantive principles that guide-action for the constituents of a realistic utopia.

There are three parts to this chapter. First, I defend the division of theoretical approaches according to their view on the function of a theory of justice. I defend this approach against the numerous alternatives discussed in the ideal/non-ideal theory literature. Second, I outline the RU approach in more detail. In doing so, I clarify the RU interpretation of action-guidingness and outline the essential nature of the methodology imposed by such an interpretation (idealisations limited by feasibility constraints). Outlining the RU approach also gives me the opportunity to defend my focus on it by elaborating on its value. In part three, I demonstrate the prevalence in the literature of both the aim of RU conceptions of justice and the broad methodological commitments it entails. I clarify the main disagreements between RU theorists when it comes to a more substantial conception of feasibility. Discussing these disagreements allows me to introduce the three requirements that I argue in the rest of the thesis RU theorists should adopt. Theorists should adopt them because they draw on the core commitment to producing a conception of justice that governs the constituents of a moderately idealised utopia. Acknowledging such requirements gives us a fully worked-out definition of the RU approach.
1.1 Dividing Methodologies by Their Aim

Political theories differ with regards to not only the substantive principles of justice (its conception) but also the way in which these principles are produced (its approach or methodology). Because of the diversity of approaches employed by various theorists, it’s helpful to categorise them according to shared justificatory and methodological principles. If we establish such categories, then we may be able to explain disagreements over substantive principles as the result of using a different approach, rather than a dispute over their fundamental value. For example, the approach theorists take to develop and justify the substantive claim ‘states should distribute wealth equally’ differs depending on whether they intend that claim to apply to states here and now, in some nearby possible world with some moderately idealised features, or in some distant possible world with highly idealised features. The intended application of a conception will affect how theorists justify it and the method they use to develop it. If they intend the conception to apply immediately, here and now, then the one with the least implementation costs might be the most justified. If proposing a conception of justice for an ideal world, on the other hand, the practicality of implementation might not contribute to its justification at all. As such, the source of a disagreement over the conception’s soundness could be down to a methodological disagreement rather than a substantive one.

In this section I briefly summarise the nature of some of these categorisations and introduce the one that I favour. I share the view of Alan Hamlin and Zofia Stemplowska (2012) that we should categorise theories according to their function. I supplement their account with the idea that a theorist’s particular aim (within a function) can establish a second layer of categories. We shall see that each view of the function and aim of a theory of justice informs its justificatory and methodological principles. Recall that justificatory principles specify what we ought to do, and methodological principles refine the actual approach that we take to ensure that the justificatory requirements are met.

A popular and well-known property that distinguishes theoretical approaches is whether or not they are ‘ideal’. Although the ideal/non-ideal distinction is
widely recognised, there’s no consensus over the best way to specify it. Hamlin and Stemplowska identify four broad specifications other theorists adopt as i) full compliance and non-compliance, ii) idealisation and abstraction, iii) fact-sensitivity and fact-insensitivity, and iv) perfect justice and local improvements in justice (ibid., pp. 48-9). Methodologies can vary according to each of these properties and, as Hamlin and Stemplowska argue, the variations are scalar rather than binary. Each variation is a matter of degree. Binary categorisations fail because it’s unclear, for example, whether a theory developed under the assumption of partial compliance is sufficiently ideal to be included in the category of ideal theory or sufficiently realistic to be regarded as non-ideal theory. It’s not plausible that once the extent of idealisations passes a certain threshold that a theorist’s task suddenly belongs to a categorically distinct methodology. Would the threshold stipulate that considering any single individual’s non-compliance equates to non-ideal theory? Does the basis of non-compliance matter? What if, generally, everyone complies but each person violates one rule? ‘Given the presence of at least two variables—the number of compliers and the extent of compliance by each—non-full compliance can take a number of forms, giving us a continuum of (non-)compliance’ (ibid., p. 49).

We should be interested in the degree of idealisation involved in particular partial compliance theories, rather than trying to establish that they all belong to one category or another. Similar difficulties present themselves for the other binary specifications, too.

In addition to this mistake of applying binary categorisations when they are in-fact scalar, ‘definitions of the ideal/non-ideal distinction that focus on one, or a small number, of a possible set of relevant dimensions may obscure other dimensions and so distort normative theory’ (ibid., p. 52). Focusing on a single (or small number of) properties when categorising normative theory distorts them because one particular property does not represent the fundamental difference between ideal and non-ideal theory. Instead, we should adopt an account that can incorporate variations across multiple properties. Hamlin and Stemplowska suggest such an account and describe it as a ‘multidimensional continuum’ of ideal/non-ideal theory. This allows them to avoid making false categorical distinctions between properties, and to allow differences in theories
to be attributable to a variation in more than one property (or ‘dimension’). The list of properties includes (but is not limited to) those specifications that others in the literature have adopted: compliance, idealisation (as distinct from abstraction), fact-sensitivity, and local or universal improvements in justice.

Hamlin and Stemplowska distinguish approaches that lie on the ideal/non-ideal continuum from those that pursue the ‘Theory of Ideals’. This distinction is based on a categorical difference in the function of the theory. The function of substantive principles developed according to an approach on the ideal/non-ideal continuum is to develop a conception of justice that can regulate ‘social arrangements that will promote, instantiate, honour or otherwise deliver on the relevant ideals’ (ibid.), whilst the function of substantive principles developed according to the Theory of Ideals is to ‘identify, elucidate and clarify the nature of an idea or ideals’ (ibid.). This, I think, is the correct way to distinguish the multitude of theoretical approaches to justice; these two views of the function of a theory of justice are categorically distinct. It’s plausible for both kinds of theorists to take the purpose of political theory to be to ‘produce the right principles of justice’ and to maintain this distinction. The difference is that proponents of the Theory of Ideals think that reflecting on ideals is the right way to do this, whereas theorists on the continuum think that some kind of speculation about social arrangements is necessary to this end. Unlike the other properties that Hamlin and Stemplowska identify, the function of a theory to either be something that regulates social arrangements, or something that provides an interpretation of the concept of justice (and its related ideals) is a binary, rather than scalar, property.

As the diagram overleaf indicates, approaches along the continuum reflect the view that the function of substantive principles of justice is to regulate social arrangements.

---

4 Developing social arrangements might be taken to mean institutional design, as Hamlin and Stemplowska describe it, or monist or pluralist ideals that guide the behaviour of individuals. Although it is an interesting question which of these principles of justice ought to regulate, I leave this question aside and use ‘social arrangements’ as a placeholder appropriate to refer to either.

5 To be clear, I don’t wish to commit to any particular motivations for adopting one type over the other. Theorists may adopt whichever they view to be the most appropriate for their particular end and defend it as such.
arrangements. Many theorists refer to this function as providing ‘action-guiding recommendations’. This interpretation leads theorists to design substantive conceptions of justice with a concept of a social world in mind. An approach is action-guiding in the sense that its function is to provide recommendations on how to design social arrangements for all theorists on the continuum. I want to supplement Hamlin and Stemplowska’s analysis by distinguishing ranges of approaches within the continuum according to the background characteristics of the social world in which substantive principles are intended to perform their regulative function. This is to distinguish theories according to their approach, rather than according to any substantive claims theorists make about the suitability of one method over others to develop a conception of justice. There are three noteworthy perspectives. Some aim to produce substantive principles for a real society, others a moderately idealised realistic utopia, and others a fully utopian society. Variations between the properties on the multidimensional continuum (compliance, fact-sensitivity, etc.) are possible within each type of approach, but are restricted by their assumptions about the background characteristics of the social world and the constituency they intend their substantive principles to govern. Each approach represents a narrowed range of the multidimensional continuum, and does not refer to a particular worked-out methodology.

![Diagram](image)

**Figure 2. Functions and Aims of Conceptions of Justice**

The commitment to either function of substantive principles of justice, and particular aim within the continuum, informs the methodological and

---

6 So, some theorists might think their preferred approach is the only way to use political theory to regulate social arrangements. Adopting the preferred approach does not, however, entail this value claim. Distinguishing approaches in this way is compatible with any number of claims about the value of particular approaches for any particular ends.
justificatory principles of the approach. The three approaches within the continuum interpret the social world to which principles of justice are to apply differently. Non-Idealist (NI) theorists consider a particular society that exists in the world today. In this sense the substantive principles they produce must be immediately implementable. Realistic Utopian (RU) theorists, on the other hand, idealise the society to which their substantive principles are to apply although the extent of these idealisations is limited by what is feasible. So, some RU theorist might consider a world that resembles ours, but could only be possible in years to come given some possible but demanding changes. Their conceptions of justice are limited by the requirement that they be ultimately implementable. Utopian theorists reject the limitation to realistic societies and instead design substantive principles for a world that doesn’t necessarily share features of our current world. As such, they don’t consider the implementability of their substantive principles at all.

Although the features of the social world for which they design a conception of justice are categorically distinct, this supplement is compatible with Hamlin and Stemplowska’s analysis. It distinguishes theories according to their aim, and permits variations along multiple dimensions of the continuum within the constraints of each approach. The methodological constraints of each approach are given by their view of the social world for which they are designing action-guiding recommendations. They essentially reduce to a difference in the way that theorists use idealisations.

The use of idealisations captures a number of dimensions that vary along the ideal/non-ideal continuum. When a theorist uses an idealisation it means that she assumes some falsehood about the empirical nature of the situation for which she is prescribing substantive principles. So, the theorist replaces ‘empirically substantiated information about persons, society, and institutions with certain idealizing assumptions, and then proceeds by determining which

\[ \text{\footnote{I will, for the most part, discuss the feasibility of a conception of justice. A conception of justice is feasible when its implementation is so. This means that we have to consider institutional requirements of the approach, the empirical circumstances facing the society to which it will be implemented, and the citizens to whom it is intended to guide. I say much more about these things as the chapter progresses.}} \]
principles of justice would be appropriate for a society in which these assumptions held’ (Weinberg, 2013, p. 331). If a theorist is developing substantive principles for a Utopian society, for instance, then she may idealise any features of our empirical world. RU theorists, on the other hand, must constrain their idealisations so that their substantive principles could eventually be implemented in the world as we know it. They could idealise away something like the current fiscal constraints of a society, but not something like citizens’ need to eat every day. NI theorists don’t use idealisations at all. By highlighting the use of idealisations in my discussion, it is not my intention reject Hamlin and Stemplowska’s analysis by arguing them to be the single property that distinguishes approaches. They reject idealisation and abstraction (or the lack of idealisation) as a binary categorisation of theories (2012, pp. 50-1). Their rejection is based on a mistake in the distinction between the two concepts. I say more about this shortly but for now merely want to distinguish the sense in which I use the term ‘idealisation’. Rather than considering it to represent one property on the continuum, I intend it to capture any shift in one or more of the variables. A presumption about compliance, fact-sensitivity, or the scope of justice can all be understood as an idealisation in some sense. It is still useful to consider the continuum as multidimensional so that the details and nature of idealisations can be analysed in more detail.

With this in mind, it is important to consider briefly the supposed distinction between idealisation and abstraction, since both (if they are distinct) could plausibly be said to feature as a property within the ideal/non-ideal continuum. Some theorists describe the utopian element of their approach as abstracting from facts, rather than idealising them. Onora O’Neill, a famous proponent of the distinction between idealisation and abstraction, describes the latter as ‘a matter of bracketing, but not denying, predicates that are true of the matter under discussion’ (O’Neill, 1996, p. 40). Idealisation, on the other hand, ‘ascribes predicates that are false of the case in hand, and so often denies predicates that are true of that case’ (ibid., p. 41). Some theorists, including O’Neill, object to the latter on the basis of this distinction. In the following short discussion, I explain why I think that this distinction is flawed and thus
why I assume that using an idealisation does not undermine an approach to justice in the way that she is concerned.

The distinction is flawed because it’s unclear that the two activities are relevantly distinct. As Miriam Ronzoni observes, O’Neill ‘claims that abstraction only consists of leaving aside predicates of certain objects that are not relevant to the matter under discussion: but what, if not a certain ideal or guideline, can tell us what is indeed relevant and what is not?’ (2010, p. 55). So, bracketing predicates is not relevantly different from denying them since either action requires an appeal to some ideal or other type of justification. Ronzoni uses the example of perfect rationality to elucidate this point. It might seem as though assuming perfect rationality is an idealisation of the capacities of real people since it is false that all real people are perfectly rational and we’re appealing to an ideal if we say that they are. However, we could also interpret perfect rationality as employing an abstraction since we could say that we are simply bracketing those predicates that make human rationality imperfect, e.g. the fact that people are inconsistent or that they sometimes cannot stick to their plans (ibid., p. 92). Whilst bracketing those predicates, we are employing an implicit assumption about which should and shouldn’t be bracketed and thus employing a similar ideal as we do when we idealise the facts. In this sense, idealisation can be reduced to abstraction and supposed difficulties with the former cannot necessarily be avoided by only using latter. Whether or not the distinction holds is not the focus of the chapter. I take this short analysis to be sufficient to show that the forthcoming discussion is independent of the way that theorists describe the utopian aspect of their approach. My use of the term ‘idealisation’ isn’t intended to exclude those who take themselves to employ abstractions only. When I talk about idealisations I intend to refer to both those theorists who use idealisations and those who use abstractions.

In the discussion that follows I outline the nature of each approach in more detail and give legitimacy to the criticisms that face NI and Utopian approaches. Doing so doesn’t conclusively defend the priority of RU, but draws support for the claim that theorists should see it as a viable option.
1.1.1 Non-Ideal (NI) Approaches

As we have seen, all theories that lie on the ideal/non-ideal continuum produce substantive principles of justice that regulate social arrangements or, in other words, provide action-guiding recommendations. There are three ranges of approaches on the continuum, each of which interprets the aim to produce action-guiding recommendations differently, via a particular understanding of the constituency of people whom substantive principles will guide. Variation across all of the dimensions on the continuum is given by theorists' different assumptions about the background characteristics of the world to which their conception of justice applies. Theorists who pursue Non-Ideal theory, a particular range of approaches on the continuum, think that conceptions of justice ought to provide action-guiding recommendations for here and now. On their view, the citizens of today are the appropriate constituency. For these theorists, a principle is action-guiding when it describes something that actual people here and now ought to do. This limits the range of variation of dimensions on the continuum.

This interpretation, that the aim of conceptions of justice is to regulate social arrangements, is plausible. Not only is their interpretation of what it means for a principle to be action-guiding intuitive, but we can be confident that recommendations made on the basis of it will have a good effect in terms of advancing just states of affairs. The desire to make recommendations to people in the real world so that conceptions of justice can make a difference to societies we live in today is one that many theorists share and the NI interpretation of the aim of conceptions of justice reflects this ideal. NI theorists might also hold the view that 'It is only in this area, the realm of real choice in the real world, where a fairly specific range of different outcomes is possible but the precise outcome is not certain, that one can talk seriously

---

*I'm presuming that the theorist is considering here and now. It's perfectly compatible for proponents of this approach to consider an alternative place (i.e. not here), so long as it is a place that currently exists. She could, for example, consider principles of justice for Nigeria, rather than the UK. In that case the relevant constituency for guidance, i.e. those that her principles would prescribe action for, are those in Nigeria. Alternatively, the theorist could consider global principles, in which case citizens of all countries would form the relevant constituency.*
about good and bad’ (Carens, 1996, p. 160). Furthermore, since we’ve a great deal of knowledge about the world around us, both moral and non-moral, we should be able to make confident claims about what actual people in this world should do, and the likely effects of their doing so.

Aiming to provide action-guiding recommendations for people here and now distinguishes this approach from alternatives by requiring NI theorists to consider the immediate implementation of their conceptions of justice.9 When a principle must be immediately implementable, then theorists are constrained by the facts surrounding the status quo in a number of ways. This commitment equates to a justificatory claim that ‘justice cannot be impractical’ (Weinberg, 2013, p. 335) where impracticality must be considered on a relatively short-term scale. So, an NI theory of justice must include justificatory principles that require conceptions of justice to be sensitive to facts that shape the status quo since these inform whether a principle can be implemented and thus whether it is action-guiding in the NI sense. For example, a well-established totalitarian regime, such as the kind in North Korea today, might preclude the implementation of a democratic conception of justice. An NI theorist considering action-guiding substantive principles for the citizens of North Korea would have to consider the actions available to these citizens given the social and political situation they are faced with. If the actions they recommend are impossible, then the conception of justice cannot be implemented and must be revised.

Clearly, this feature of the approach constrains the way in which NI theorists may use idealisations. If considering immediate implementation requires substantive principles be sensitive to facts surrounding the status quo, then NI theorists should not idealise those facts. Doing so would undermine the resulting principles’ ability to guide action in the real world where the

9 ‘Immediate’ does not need to be taken literally here, since NI theorists could plausibly think that a principle can be action-guiding so long as it can be implemented in the near future (where near could be interpreted as a matter of days, weeks, or months). I use the phrase ‘immediate implementation’ to distinguish NI concerns from the ‘ultimate implementation’ that concerns RU theorists and I discuss in the following section.
idealisations no-longer obtain. Since guidance is intended for actual people here and now, idealisations are not helpful or necessary for this approach.

1.1.2 Realistic Utopian (RU) Approaches

Realistic Utopianism is an alternative group of approaches on the ideal/non-ideal continuum. Like the non-idealists, these theorists view the function of a theory of justice to be to produce action-guiding recommendations. However, they interpret the meaning of action-guiding differently to NI theorists so that they aim to produce substantive principles that guide the constituents of a moderately idealised utopia. This interpretation of action-guiding recommendations, and understanding of the aim of a conception of justice to be to produce them, is what I call the ‘core commitment’ of the RU approach. Rather than, like NI theorists, asking whether a principle is immediately implementable, RU theorists are interested in whether a conception is ‘ultimately implementable’ (Weinberg, 2013, p. 335). This idea of an ultimately implementable conception of justice is crucial and definitive of the RU approach. A conception is ultimately implementable when it could govern some possible version of the world as we know it. This captures the sense in which the RU approach is distinct from the NI portion of the continuum—by dropping the idea that conceptions need to be immediately implementable, theorists are free to idealise certain aspects of the society in which they intend substantive principles to apply. This describes the utopian element of the approach. The range of idealisations compatible with it is constrained by the idea that the utopia is only moderately idealised, or in some sense realistic. The requirement to produce ultimately implementable conceptions of justice captures this aspect of the approach also. It distinguishes them from the Utopian portion of the ideal/non-ideal continuum because implementability requires feasibility to be a consideration in the formulation of substantive principles of justice, something that Utopian theorists reject.

An objection to the NI approach reveals one sense in which RU theory is appealing. The objection claims that NI theorists lack the necessary knowledge to make claims about justice even within the constraints of the immediately implementable. Although they offer reasons in favour of a proposal, they need
to understand how these reasons justify the claim that a set of arrangements are ideal under present historical circumstances. Doing so will require some understanding of what weight (if any) those considerations would have in different circumstances when other considerations also enter the picture and need to be taken into account (Mason, 2004, p. 254). So, in order to understand such reasoning, theorists need to make some more abstract considerations than the NI approach permits. Swift illustrates this problem well with the following example: ‘In an ideally just society parents would act wrongly if they purchased for their children an education superior to that provided by the state…but to know whether it is wrong for parents to purchase this kind of education, or what kind of education policy is the right one here and now, we have to consider the values that explain why those claims are justified…It is only by being clear at the philosophical level that we can work out what justice demands of us’ (Swift, 2008, p. 378). NI theorists must be able to consider an ideal world in order to make the kinds of normative claims that they aim for.

Making more abstract considerations ‘may not require us to address the question “What is ideal under optimal conceivable conditions?”, but it does require us to abstract from the particular set of circumstances which limit current possibility, otherwise there is the danger that our ideals will be inadequately rather than better grounded’ (Mason, 2004, p. 254). Given this limitation of the NI approach, we might prefer the RU alternative. I don’t take this objection to be sufficiently expressed or defended for us to reject NI theory as a plausible aim altogether, but outlining it briefly here does motivate our acceptance of the RU modification to it so that we endorse more abstract reasoning.

As this discussion demonstrates, one methodological implication of the RU interpretation of action-guiding, as opposed to the NI version, is the use of idealisations. ‘Rather than taking the feasible set as given, it might be thought that it is precisely the task of the philosopher to change that set, by changing the content of those popular beliefs that do so much to determine it. This is at the very least a widespread understanding of the proper role of the political
philosopher.’ (Swift, 1999, p. 359) As Swift says, one of the reasons theorists might produce a theory of justice is to persuade people to act more justly than they currently do. As such, theorists must consider and perhaps endorse desirable actions even though their fulfilment may not be possible in current circumstances. As widespread as this view may be, the reason particular theorists adopt the RU approach over alternatives is not at issue.\(^\text{10}\) I remain neutral regarding the question of why theorists adopt one approach over alternatives and instead establish its plausibility as one option alongside alternatives. Later, I discuss the value of the RU approach in more detail, although I do so only to give it some priority over alternatives, and not to defend the idea that it should be adopted at their expense. The focus of the thesis is to consider the methodological implications, given theorists are pursuing the RU approach.

As we have seen, the substantive principles of an RU theory of justice do not intend to guide the actions of actual people living in the world today, but rather a certain type of people (which may be different from you and I) in a certain type of world (which may be different from here and now). As such, the RU approach can avoid the criticisms of the NI approach by using idealisations. RU approaches vary according to the details of these idealisations; theorists give arguments that demonstrate why particular idealisations are appropriate. Despite this room for variation, RU theorists agree that idealisations must be limited. This limitation represents the second broad consensus within their approach. Idealisations are limited so that the conceptions produced under their assumption are ultimately implementable into the real world. A commonly accepted necessary condition of producing conceptions that are implementable in this sense is that they be feasible. So, RU theory is subject to feasibility constraints. Feasibility acts as a constraint on requirements of justice when it explains how certain empirical circumstances block particular requirements during substantive principles’ formulation. This second

\(^{10}\text{Alternatively, theorists might think that the RU approach is valuable because it informs their reasoning when deliberating what justice requires of them today, or purely out of intellectual curiosity.}\)
methodological commitment distinguishes RU theory from the remaining group on the ideal/non-ideal continuum; Utopian approaches.

So, a Realistic Utopian approach to justice uses idealisations to develop substantive principles for a moderately idealised society, but is constrained by some account of feasibility that makes those principles ultimately implementable to the real world. I discuss the core commitment to producing an ultimately implementable conception of justice and the essential methodological requirements of the approach in more detail in section two.

1.1.3 Utopian (U) Approaches

The final range of approaches on the Ideal/Non-ideal continuum I call ‘Utopian’. Proponents of this view, like NI and RU theorists, take the function of a theory of justice to be to provide action-guiding recommendations that guide social arrangements and introduce a third interpretation of this aim. The constituents to whom these theorists’ recommendations apply inhabit a utopia. For them, a principle of justice is action-guiding if it is in any sense prescriptive. This matches what Weinberg calls formal practicality. It merely describes the nature of a conception as connoting action and conceptions meet this requirement if they tell some agents to do something (2013, p. 331). Their view is that one can develop substantive principles that regulate social arrangements by considering what justice requires for the constituents of a world with, for example, super abundance or perfect altruism. Carens’ *Equality, Moral Incentives, and the Market* (1980) is a good example of this kind of approach. He takes substantive principles of justice to design a social order, but does not think it should be constrained by existing institutions or other empirical facts. Although they share the idea with RU theorists that utopian theorising is desirable, Utopian theory is distinct because these theorists reject the idea that sensitivity to feasibility constraints is a necessary condition of producing action-guiding principles.

The sense in which Utopian conceptions of justice guide action does not match very well with the ordinary usage of the phrase. It is, however, in at least one sense plausible. Substantive principles can have a prescriptive nature without being immediately, or clearly, applicable to people today. Estlund
describes one sense of this nature: ‘it counsels all people, together, to behave differently. It does not tell each person to do it whether or not others do...a hopeless normative conception of society might be collectively prescriptive in that same way without being individually prescriptive at all...So, the fact that under the circumstances the theory doesn’t recommend any particular actions by any individual doesn’t show that it doesn’t counsel action. It remains aspirational: it sets standards that are not met, but could be met, and it tells us to meet them’ (Estlund, 2008, pp. 266-7). Estlund is talking about standards which ordinary people do not meet, and may never meet (although there is nothing preventing it being possible). Utopian theorists may prescribe standards like these, but its scope is not limited by possibility.

A fundamental difference between this approach and the others discussed so far is that it does not require the theorist to consider implementing her substantive principles (in any sense of the word) whilst developing them. Because of this, the Utopian approach is not constrained by facts describing the status quo or other feasibility considerations. Some might think that it’s problematic that justice may not be something that we can ever implement in the real world. ‘If what we are interested in is guidance in this world, trying to construct a theory that aims to be valid across all possible worlds, including very remote ones, will come at serious costs. As Pogge puts it, in so doing, ‘we are likely to achieve less unity and likely also to trade off fit with our considered judgements about this world (which matters greatly) against fit with our judgements about very different worlds (which matters little)’ (Pogge, 2008, p. 475)’ (Valentini, 2011, p. 31). Utopians could, for example, recommend a system of direct democracy, arguing that justice requires all citizens to have a say in each and every important political decision. Although our intuitions might be in favour of such a system in theory (since it secures the justness/legitimacy/autonomy of political policies created in this way), in practice our societies are too big and too pluralistic for this to ever be a practical system, even in the best of foreseeable conditions. Carens describes a similar concern that: ‘the characteristic danger of the idealistic approach to morality is that such an approach may be irrelevant to the moral issues we face. Idealistic inquiries may be academic in the pejorative sense of that term,
privileged speculations that do little to help us reflect upon the moral choices we must make or to guide us to act responsibly in the world’ (Carens, 1996, p. 168). It's not clear what the success of direct democracy in an ideal world can tell us about the kinds of political system that we should favour today. What theorists who criticise ideal theory ‘really want is work that will give us concrete guidance into the particular contexts or circumstances that we actually face, and guidance that there is some prospect of our actually being guided by’ (Swift, 2008, p. 368). On the other hand, it’s not clear that we should consider the task of justice to be to tell us what we should favour today. However, given these difficulties, we might prefer the RU approach since it occupies an attractive middle-ground.

1.1.4 Theory of Ideals (ToI) Approaches

As we have seen, the Theory of Ideals refers to approaches to justice that fall outside of the ideal/non-ideal continuum. Proponents of this approach have a different view of the function of a theory of justice. Rather than develop substantive principles that guide social arrangements, these theorists’ interpret the function of a theory of justice to be to provide an accurate reflection of the truth of the ideal. 11 The reason for doing this is to ‘know or understand something about justice rather than to motivate action towards it’ (ibid., p. 366). According to proponents of this view, justice needn’t guide action since there’s no reason for us to think of it as something that ‘the state, or, indeed, any other agent, is in a position to deliver’ (Cohen, 2004, p. 18). These theorists may endorse any approach on the ideal/non-ideal continuum when it comes to designing social arrangements, but they understand this task to be distinct from their own (and therefore don’t think that the results should be understood as principles of justice).

The ToI involves taking an approach to develop conceptions of justice ‘from the viewpoint of justice alone’ (Cohen, 2008, p. 275). This shapes the approach in the sense that, like Utopians, proponents don’t take feasibility or the status

---

11 Note that theorists on the continuum might also think the results of their approach constitute the truth of the ideal. The distinction lies in the former’s insistence that the nature of a conception of justice is to guide social arrangements.
quo to be relevant concerns when developing substantive principles of justice. They are distinct, recall, because the former takes them to describe social arrangements, even though those arrangements may order some distant, abstract, society. Cohen’s work ‘gives us tools to answer the evaluative question of what counts as a just state of affairs (tools that are meant to be valid across all possible worlds), but does not directly address the normative question’ (Valentini, 2011, p. 30). The normative question requires conceptions of justice to yield action-guiding recommendations, but these do not concern proponents of ToI because ‘sometimes, justice is unattainable’ (Cohen, 2008, p. 84). Not only do these theorists not need to consider the implementation of their conceptions, but they ‘can only pursue the general inquiry into the nature and structure of values successfully if [they] are not tied to any particular feasibility constraint’ (Hamlin & Stemplowska, 2012, p. 55, emphasis added).

So, a methodological implication of this approach is that theorists must reject substantive principles that make the justifiability of a conception of justice sensitive to its empirical surroundings.

We might consider whether there is a relation between ToI approaches and those that lie on the ideal/non-ideal continuum. It could be the case, for instance, that an RU theorist views the product of ToI theorising to be a component in her arguments about what justice requires of the constituents of a moderately idealised society. This view isn’t, however, entailed, by the RU theorists’ aim to produce this kind of action-guiding recommendation. Similarly, theorists pursuing the ToI might think that some of the abstract theorising done by Utopian theorists is correct, in the sense that it successfully maps requirements of justice, but deny that considering social arrangements was necessary to produce such results or that it contributes to their justification. Again, this won’t necessarily be the case since it depends on the

---

12 An exception to this claim would be if a ToI theorist took it to be a part of the truth of the ideal of justice that it is sensitive to facts of feasibility. Although this position is coherent, it's not very plausible and not held by any existing theorists, as far as I am aware. Nevertheless, it is still distinct from approaches on the continuum since its sensitivity to feasibility is the result of a substantive commitment. For RU and NI theorists, feasibility constraints are a methodological requirement that results from their interpretation of the aim to produce action-guiding recommendations.
substantive content of the conception of justice and is not a relation entailed by the characteristics of the methodologies themselves.

Analytic political theories designed to produce conceptions of justice fall into one of the four categories discussed here. Each interpretation of the function of a theory and the aim of a conception of justice, commits proponents to certain methodological constraints, as I have outlined briefly in the preceding paragraphs. These constraints narrow the range of approaches on the ideal/non-ideal continuum. Although I have not conclusively defended the value of RU theory over its alternatives, I have indicated what I take to be the main reasons in favour of doing so. At the very least the large range of existing methodologies within its constraints is a reason to favour it. I discuss examples of these in part three of the chapter. In the next section, I develop the methodological commitments of the RU approach in more detail. Doing so allows me to expand on its value to show why we might give it priority over alternative approaches without defending the stronger claim that we ought to adopt it at their expense.

1.2 The Realistic Utopian Approach

In this section I say more about what it means to take an RU approach to developing conceptions of justice, in particular noting the implications these theorists’ interpretation of action-guiding recommendations has on the methodology they may adopt. The defining characteristic of RU theorists, as we have seen, is a commitment to developing a conception of justice to govern the constituents of a moderately idealised utopia that recommends actions that are ultimately implementable to the world as we know it. A major methodological implication of this aim is that the limits of permissible idealisations are set by feasibility constraints. Theorists cannot idealise the society to which their substantive principles apply if doing so makes the results
Feasibility must be built into the methodology so that it shapes the content of substantive principles from the start. This implication is somewhat problematic since theorists disagree over anything more substantial than a very thin and uncontroversial conception of feasibility and therefore disagree over the extent to which idealisations are appropriate.

1.2.1 RU Idealisations

As we have seen, the use of idealisations defines the RU approach and distinguishes it from the NI alternative. Particular RU approaches idealise different facts; theorists can make arguments to include particular idealisations. For example, some theorists may argue that conceptions of justice should be developed under the assumption that society is closed, so that citizens cannot enter and leave as they wish.

The RU approach is distinguished from the Utopian alternative, recall, because its aim is to produce conceptions of justice that can be ultimately implemented to the world as we know it. In this sense, idealisations are permissible so long as they do not undermine a good-faith interpretation of what it means to produce a conception of justice for a moderately idealised utopia.

The idea that RU theorists must constrain their use of idealisations is widely embraced in the literature. If substantive principles fail to guide action, then they fail to be effective principles of justice. This claim is known as the ‘guidance critique’ (Valentini, 2009, p. 333) and is compatible with the way in which RU theorists understand constraints on idealisations. Valentini’s condition on idealisation states that: ‘the appeal to ideal predicates in the design of a theory of justice does not inevitably lead to unsound practical conclusions, at least so long as such predicates are not assumed as a condition for the applicability of the theory’ (ibid., p. 353). When Valentini demands ‘practical conclusions’, she means guidance that applies in the real world. As we know, this is not a requirement of the RU approach—their guidance is intended for the constituents of a utopia—but the intuition behind it matches

---

13 Whether or not current incompatibility with the world here and now makes a principle infeasible requires more attention and is something that I consider in the following subsection on feasibility.
their view that conceptions should be ultimately implementable. RU theorists take a good-faith interpretation of action-guiding to mean that it must be possible for their conceptions to ultimately be implemented, but not that they must be immediately implementable into society today (as is the view of NI theorists). As such, they agree with Valentini in the sense that the idealised circumstances in which conceptions are developed cannot undermine their normative force in a feasible version of the world as we know it. For Valentini, this means that principles must still apply when the idealised conditions no-longer obtain. For RU theorists, it means that idealised conditions must be constrained so that they do not describe a world that is too dissimilar from our own. Work is required to establish what kind of world is ‘too dissimilar’.

This constraint on the use of idealisations isn’t to say that the label ‘RU’ no-longer reflects a range of approaches within the ideal/non-ideal continuum. It doesn’t give determinate methodological guidelines—it doesn’t say which idealisations RU theorists should adopt—but it does affect the compatibility of methodological and justificatory principles. RU theorists must, for instance, avoid methodological principles that require conceptions of justice to be developed in wildly abstract scenarios. To avoid the guidance critique, and stay loyal to their interpretation of action-guiding recommendations, RU theory must recommend actions for beings relevantly like us in a world relevantly similar to ours. By considering this grounded kind of utopia, theorists can ensure their recommendations are ultimately implementable, as per their aim.

The focus of the thesis is to consider the methodological implications of making a conception of justice implementable in this sense. One such implication is acknowledging feasibility constraints.

1.2.2 Feasibility

Feasibility is a necessary condition of a conception being ultimately implementable. Substantive principles must be feasible in order to guide-action in the RU sense, even though their conceptions of justice govern the constituents of an idealised society. In this section I start by recalling the link

---

14 In the remaining chapters of the thesis I defend additional necessary conditions of ultimate implementability that are not as widely shared as feasibility constraints.
between ultimate implementation and feasibility. I then establish the minimal account (i.e. thinnest) of feasibility required by the RU approach.

Realistic Utopian theorists design substantive principles aimed to guide the constituents of a moderately idealised society. This reflects their understanding of what makes principles action-guiding—they must be ultimately implementable. This interpretation of action-guidingness requires principles be substantively, rather than merely formally, practical. Substantive practicality means that it is in some sense possible for agents to comply with that recommendation (Weinberg, 2013, p. 331). This sense of practicality is shared by NI theorists in their interpretation of action-guiding recommendations, too.

To determine whether a proposal is substantively practical, and can ultimately be implemented, RU theorists must establish which facts about the real world are relevant and as such cannot be idealised. We’ve seen that one condition of ultimate implementability, and one way to establish these facts, is to make sure that conceptions of justice are feasible. However the details of a feasibility constraint remain unclear.

I have said that RU theorists must acknowledge the feasibility of conceptions of justice because doing so is necessary for them to stay loyal to their claim that conceptions of justice should be ultimately implementable. There are, of course, other reasons that we might think that feasibility is important, though. Carens quotes Hoffmann (1981, p. 28) as capturing the fundamental impulse behind the idea that conceptions of justice should be sensitive to feasibility:

One of the key necessities in this field is to avoid too big a gap between what is and what ought to be. In any system of law, or in any system of morals, there is always a gap between the is and the ought, between the empirical pattern and the norm. The gap is necessary and inevitable. If there were no gap, people would not feel any sense of obligation, or any remorse

---

15 We can evaluate the feasibility of a number of things: individual actions, policies, institutions, or states of the world to which any of the former might give rise; outcomes the policy recommends; or events. I’ll normally talk about the feasibility of implementing or adhering to conceptions of justice although what I say about the concept of feasibility does not apply exclusively to this kind of concern.
when they violate a norm. But when the gap becomes too big, the system of law or the system of morals is really doomed - to have no impact whatsoever or to be destroyed (Carens, 1996, p. 157)

Here the idea is that in order to have an impact and in a plausible sense guide us, conceptions of justice ought to be achievable. Valentini (2011) describes political theory that is sensitive to facts about the world as the ‘embedded approach to justice’. Proponents of this view think that empirical concerns are important because justice ‘does not transcend the circumstances of human social existence but is inherently tied to them. Moreover, given that, for proponents of the embedded approach, principles of justice play a social function, their design must be informed not only by normative but also by empirical assumptions about the particular social system they are meant to regulate’ (Valentini, 2011, p. 27). This reflects a similar idea to Carens’ and Hoffman’s. Because these theorists want conceptions of justice to address social worlds as we find them or plausibly could find them, we should regard them as sensitive to the world in which people act.

Some theorists disagree that feasibility should affect the content of a conception of justice in this way. On this view, whether or not an action is feasible is completely independent to the question of whether it is just. It follows from this view that some circumstances are inherently unjust even though there’s nothing we can feasibly do to change their regrettable nature. Gheaus (2013) uses the example of a natural disaster to make this point. Surely, she says, inequalities that result from the effects of a hurricane are exactly the kinds of circumstances that we think of as unjust. Even though it wasn’t possible, or feasible, for us to do anything to avoid these affects, the resulting circumstances are still unjust. It seems to me, however, that it’s perfectly in-line with our intuitions to describe the effects of a hurricane as unfortunate, without making the claim that they constitute an injustice in virtue of the fact that some are affected more than others. It’s not until resources are allocated

16 I say more about the relation between possibility and feasibility later in the chapter.
to fix the damage that our intuitions about justice kick-in. If, after the hurricane, every person affected received £1,000 of aid, but for some this is more than adequate and for others not nearly enough, then we invoke principles of justice to either defend or dispute the distribution. My response to Gheaus resembles Swift's view when he says that we 'misdescribe a state of affairs as unjust if we identify it as being bad in ways that we might regret but that human agents—collectively or individually—are incapable of remedying. We can make no claim about the moral badness—they are simply unfortunate facts about the world—and they cannot be regarded as unjust' (Swift, 2008, p. 367). Humans cannot do anything about the damage an unforeseeable hurricane inflicts. What we can do is devise rules about whether and how we should respond to the effects such damage has on people's lives. So, it makes perfect sense to think that the former does not trigger our intuitions about justice whereas the latter does. It's worth noting here that the view that we should treat feasibility as entirely separate to justice is often paired with a dispute over the RU interpretation of action-guiding. Their interpretation, recall, says that the aim of a conception of justice is to govern the constituents of a moderately idealised utopia. The fact that theorists developing Theory of Ideals or Utopian theory disregard the link between justice and feasibility is of no concern. I've defended the claim that RU conceptions of justice ought to be feasibly implementable, but our intuitions on the plausibility of this claim will, of course, depend on the particular conception of feasibility.

The precise features of a feasible conception of justice are difficult to identify and defend. It is clear from the presence of idealisations in the approach that the relevant sense of feasibility for RU theorists isn’t so restrictive that it confines them to the precise conditions of the world as we know it. So, to be feasible, it doesn’t have to be true that theorists could immediately implement a conception of justice in the empirical circumstances of here and now.

The problem here is that the limits of the possible are not given by the actual, for we can to a greater or lesser extent change political and social institutions and much else. Hence we have to rely on conjecture and speculation, arguing as best we can
that the social world we envision is feasible and might actually exist, if not now then at some future time under happier circumstances (1999, p. 12).

To be feasible, a conception might only need to be possible in a version of the world somewhat distant to the one we are familiar with now. I leave aside the epistemic question of what’s necessary to say that we know something like this. Even after refining themselves to the limits of the possible rather than the actual, theorists work with different conceptions of feasibility and thus different ideas of which facts may inform their conception of justice. Although there is scope for different conceptions of feasibility within the range of approaches on the RU portion of the continuum, their core commitment establishes some limitations on compatible conceptions. In the following analysis of feasibility, I am interested in establishing an account that is plausible, expresses the minimum requirements that follow from the core commitment, and is loyal to existing use. We’ll see that RU theorists’ idealised societies share our laws of science, are subject to the circumstances of justice, and facts that inform the reasonable likelihood of success. These features establish a minimal conception of feasibility required by the core commitment of the approach. A more substantial conception of feasibility may be compatible but, as I shall discuss at the end of the chapter, is not required as an essential part of the RU approach.

1.2.2.1 An Uncontroversial Conception of Feasibility

The principle ‘ought implies can’ is widely accepted and reflects many theorists’ view that compliance with conceptions of justice should at least be possible (otherwise they fail to qualify as requirements of justice). If you have slipped and are falling from a cliff to certain death, and the only way that I can save you is to turn-back time and stop you from slipping, then justice cannot require that I save you. The sole reason that justice permits my failing to save you (assuming that normally it would require me to do so) is because I can’t.

17 Hamlin and Stemplowska describe this as a tension between the imaginable and the truly reachable (2012, p. 17). Although this epistemic dimension raises important and interesting questions, addressing it is beyond the scope of the thesis.
This example reflects a very thin conception of feasibility that all RU theorists accept (actions that it is and always will be impossible for humans to perform are excluded as putative requirements of justice). Unless we adopt an unusual conception of possibility, the only actions that this principle precludes as requirements of justice are those that are impossible according to the ‘true laws of science’. Rawls includes these facts, what he calls the ‘actual laws of nature’, into his conception of a realistic utopia (Rawls, 1999, p. 13). Hamlin and Stemplowska describe this as the ‘possible worlds’ conception of feasibility (2012, p. 56). On this view, ‘if an alternative social arrangement or an alternative account of the motivation of individuals is possible in the purely technical sense, then it is included in the relevant feasible set’ (ibid.). I refer to this as the possible worlds (PW) account of feasibility. This conception does not make justice sensitive to all the facts that describe our circumstances now, only those that are so permanent and universal that they will always do so. So things that violate the laws of physics—time travel for instance—are unfeasible, but others that seem unlikely or demanding in some other sense—like giving away all of your money to charity—are feasible.

The PW account, although plausible, is too thin to loyally represent the way in which existing RU theorists use the concept of feasibility. It’s important for me to describe these broad commitments of the approach accurately to demonstrate that there is a shared consensus and that the RU label has substance. Moreover this account of feasibility is made implausible by its delivery of counter-intuitive results in numerous types of cases. I discuss some examples of these shortly. As such, I must supplement it with a more substantive account.

For many, a conception must be sensitive to the circumstances of justice in order to be considered feasible. The circumstances of justice describe the empirical circumstances that conceptions of justice must respond to. These empirical circumstances shape our lives in such a way that make substantive principles of justice a necessary and worthwhile pursuit. They are ‘the normal

---

18 I give particular examples of existing theorists and their perspectives in the final section of the chapter.
Human cooperation is plausible because we are sufficiently similar to one another and necessary because it means that we can thrive in ways that would be impossible without cooperation. These conditions obtain ‘whenever persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity’ (ibid., p. 110). Individuals have conflicting needs and desires, and have a moderate preference for their own to be satisfied. As well as this conflict, the scarcity of goods means that each person cannot secure as much as she wants simultaneously to others doing so, too.

These conditions, first outlined and specified by Hume in *A Treatise of Human Nature* and *An Enquiry Concerning the Principles of Morals* (ibid., p. 109) are accepted almost unanimously by contemporary political theorists on the ideal/non-ideal continuum. These theorists see ‘justice as matter of conferring rights on individuals competing over scarce resources, given relevant natural and social facts… Although the question of which facts are relevant is a matter of debates among advocates of the embedded approach…they all understand justice as offering a response to normative problems generated by those facts’ (Valentini, 2011, p. 25, emphasis in original). Theorists treat these circumstances as constraints on justice when they understand the purpose of justice to be to respond to them. This means that substantive principles must be suitable for a world constrained by these circumstances since acting according to their recommendations in an alternative world would not achieve justice. A principle would be unsuitable in this way if, for example, it were to govern a world in which we all have and pursue the same interests. Similarly, a principle that prescribes a hearty distribution of natural resources for everyone may not be suitable for a world with scarce resources. These principles are unsuitable because they are infeasible for the world they are intended to guide. Conceiving of feasibility as constraints imposed by the circumstances of justice makes RU conceptions sensitive to more facts than the PW account alone. It

---

19 With the plausible exception of some within the ‘Utopian’ range of approaches on the ideal/non-ideal continuum.
makes conceptions that are incompatible with the circumstances of justice infeasible. Unlimited altruism, for example, is incompatible with the circumstances of justice, but does not conflict with a true law of science (and thus the PW account of feasibility). Intuitively, it is clear that the fact that it is technically possible for a human to be unlimitedly altruistic is not sufficient to say that such behaviour is feasible.

Finally, likelihood of success features in any intuitively plausible conception of feasibility. If probability is ignored then justice may require actions that are possible in a technical sense but very unlikely. For example, there are no true laws of science (or circumstances of justice) that prevent a ‘medical ignoramus successfully performing a neurological operation for which he lacks the relevant expertise’ (Brennan & Southwood, 2007, p. 8), yet our intuitions definitely say that such an action is unfeasible and thus the requirement that he successfully complete the action ludicrous. In this example the reason that we think that it’s unfeasible is because success in performing the specified action is extremely unlikely.

It’s clear, then, that a plausible account of feasibility must include a condition that makes successful completion of the proposed action reasonably likely. Brennan and Southwood (ibid.) suggest a conception of feasibility that accounts both for this and for some of the reasons that success is unlikely. It’s important to include these reasons because some are relevant to feasibility assessments and others are not. For Brennan and Southwood, an action is feasible if there is a reasonable probability of success conditional upon trying. Whilst the medical ignoramus’s successful performance of the operation is infeasible, a lazy father’s attendance at his daughter’s hockey match on a Saturday morning is not (ibid., p. 9), even though both might be similarly unlikely. Making the account of feasibility conditional upon trying accounts for the difference in our intuitions in these cases and makes it sensitive to the reasons for the low probability of success. Like Brennan and Southwood (ibid.), and Gilabert and Lawford-Smith (2012), I agree that a plausible conception of feasibility must be sensitive to probability. I don’t specify what the threshold of likeliness ought to be, or even whether a threshold is the best
way to account for this sensitivity, but note that likelihood must be included in
the conception of feasibility to which all RU theorists are committed (in order
for obvious cases like the medical ignoramus to be ruled out). Although
conceiving of feasibility as subject to the circumstances of justice, the laws of
science, and a condition of likelihood narrows the range of feasible actions, a
number are still up for dispute. I call the combined features of feasibility that I
have established so far the ‘uncontroversial conception’ of feasibility.

The uncontroversial conception of feasibility should be accepted by all RU
theorists, although it is not exclusive to their account. NI theorists could also
plausibly adopt it. In some cases RU theorists may regard this conception of
feasibility to be insufficient and adopt a more substantive alternative. This is
compatible with the RU approach; there’s space on the continuum for
theorists to include other factors as feasibility concerns if they see them as
appropriate, but is not required by the core commitment to producing a
conception of justice for a moderately idealised utopia. In the final section of
the chapter, I introduce three types of empirical facts that theorists might think
their conceptions ought to be sensitive to in this way. I note their intuitive
plausibility, but conclude that sensitivity to each of them originates from a
concern other than feasibility. Investigating these possibilities allows me to
introduce three commitments to the RU approach that are more stringent than
existing proponents typically acknowledge. These methodological requirements
follow from the core commitment to producing an ultimately implementable
conception of justice and reveal that although feasibility is a necessary
condition of this, it is not sufficient.

1.2.3 Value
Now that I’ve said more about RU theorists’ aim and established the broad
details of their approach, I can elaborate on its value. As I have said, there are
legitimate criticisms to RU theory’s alternatives, which give us at least a prima
facie reason to prefer it. I take those objections, plus what I am about to say
here, to be sufficient to make the case that the RU approach is worthwhile and
thus the contribution of the thesis valuable.
For some theorists, the purpose of political theory is to make the world that we’re in more just. They take developing conceptions of justice to contribute to that aim in some way. Although it is intuitively appealing, and the idea that RU theory in particular can help with such an end plausible, it’s also somewhat controversial and makes the worth of the approach merely instrumental. I want here to say a little more about the intrinsic value of RU theory. RU theorists are looking for guidance on how, within the limits of practical possibility, we can be just. As Rawls puts it: ‘our hope for the future of society rests on the belief that the social world allows at least a decent political order’ (2001, p. 4).

Although RU conceptions of justice do not tell us here and now what we should do (they’re not required to be immediately implementable), they tell us what we’re capable of doing and give us something to aim for when striving to be just. I take for granted that knowing this is intrinsically valuable. So, not only is RU theory a necessary step to deciding on conceptions of justice in the non-ideal world, but it’s also a valuable intellectual end in itself. This is not to say that understanding the requirements of justice beyond the feasible frontier is worthless, but only that it’s a separate question from the one of what we could do to be just as we can be. I’m neutral amongst theorists’ particular reasons for adopting the RU approach over alternatives and only want to clarify some of the most plausible reasons that they might.

In the literature we can see that many theorists share the view that this kind of theorising is valuable, either intrinsically or as a means to less ideal action-guidance. Dworkin favours the RU approach when he describes general philosophy as starting ‘in our life and experience’ (2000, p. 4). He takes this starting place to be beneficial because ‘only then is it likely to have the right shape, not only finally to help us, but also finally to satisfy us that the problems we have followed into the clouds are, even intellectually, genuine not spurious’ (ibid.). In the following section, I say more about the prevalence of RU theory, which I take to further support its value. Some theorists, although recognising the value of RU theorising, disagree that we ought to call the product of this approach a principle of justice. Even Cohen, a staunch and most famous advocate of the ToI, thinks that it’s important for us to consider ‘rules of
regulation’. He just takes this to be a separate task to developing a conception of justice (2008, p. 253).

I’ve not claimed that theorists should necessarily pursue RU theory at the expense of alternatives. Clearly, the immediate requirements of justice (the focus of the NI approach) are something that political theorists should also be interested in. Its value is, however, distinct from the value of RU guidance. This is because NI guidance is limited to prescriptions within the limits of immediate possibility. One might argue that the fact that NI prescriptions are immediately implementable might mean that we should pursue them at the expense of RU theory. But as we saw, it’s difficult to see how NI theorists might achieve any critical insight without at least drawing on the results of more abstract theorising.

Whether it’s because RU theory informs the reasoning that supports the claims that justice makes on citizens here and now, or whether it’s because knowing how just we could possibly be is valuable, taking the RU approach to justice is an important task for political theorists. This importance, we shall now see, is also reflected by its prevalence in the literature.

1.3 Existing Realistic Utopians

The claims in the rest of the thesis apply to any theorist who accepts the core commitment of RU theory: that the aim of a conception of justice is to provide action-guiding recommendations for the constituents of a moderately idealised utopia. I take any theorist who accepts this claim to be adopting the RU approach. In this section, I demonstrate that the core commitment of the RU approach is widely shared by existing prominent theorists. Following this, I set-out three distinct types of facts (contextual, psychological, and political) in order to investigate the areas of disagreement when it comes to a more substantial conception of feasibility than the uncontroversial account. Each introduces an empirical concern that our intuitions tell us RU substantive principles ought to be sensitive to in at least some cases, but that existing
theorists disagree over. Discussion of each of these introduces the additional methodological constraints to RU theory defended in the chapters that follow.

1.3.1 The ‘Core Commitment’ and Existing RU Theorists

The core commitment of RU theorising is widely shared by numerous influential theorists. Theorists might take this approach because they think that the results inform claims about what real people in the world today ought to do, because they see it as a valuable intellectual activity in and of itself, or for some other reason. Those that take this view include Dworkin, Miller, Nagel, Rawls, and Swift. Each makes it clear that his interest is in theorising about justice to guide action for a moderately idealised society rather than a more abstract utopia, or the immediate circumstances we are faced with. Miller says that this means that the aim of a conception of justice is to ‘guide action, rather than have a merely speculative purpose’ (2013, p. 34) and Swift makes this aim a priority by saying that ‘whatever else it may also do, political philosophy should contribute to the guidance of action’ (2008, p. 366ff). To illustrate the possible range within the RU approach I will show how two well-established and familiar methodologies are compatible. These are Rawls’s (1999) Original Position, and Dworkin’s (2002) Hypothetical Insurance Market. I note differences or compatibility between these and the more stringent RU methodological commitments I defend throughout the thesis where appropriate.

Rawls’s original position states that the soundest principles of justice are those that ‘free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association’ (1999, p. 10). Famously, these persons are placed in a hypothetical scenario—the ‘original position’—and then asked to consider fair principles based on their rational preferences. The substantive principles that result are just because the conditions of choice in the OP are fair. The information parties have in the original position reflects the way that Rawls acknowledges feasibility constraints in his method. They have information regarding general beliefs about human nature and how society works (1990, p. 565). A conception of justice, on this view, is not independent of these beliefs.
We should also note that the purpose of Rawls's hypothetical device is to design a conception of justice that governs a well-ordered society, rather than an existing one. Well-ordered societies are idealised so that the conception of justice is accepted by everyone, orders the main political and social institutions, and gives everyone reasons to act from their sense of justice so that they comply with these institutions (Rawls, 2005, p. 35). The fact that Rawls's method is designed to produce guidance for such societies, and makes substantive principles sensitive to certain empirical facts, demonstrates that he accepts the core commitment of the approach and the broad commitments of the methodology established thus far. Rawls adopts this approach over alternatives because he takes RU theory—in particular a deep understanding of ideal principles of justice—to be fundamental in the eventual allocation of duties in the real world, with all its real problems (1999, p. 8).

Dworkin’s hypothetical insurance market ‘asks a community to imagine what protection against risk, in the form of insurance, average people of normal prudence in that community would likely have purchased if they had had the opportunity to do so on equal terms, given the premiums that insurers in a competitive insurance market would have charged for such insurance’ (2002, p. 108). Distributive justice, Dworkin’s Equality of Resources, is structured according to the results of this hypothetical market. Dworkin acknowledges the function of justice to be to guide social arrangements when he says that a theory ought to ‘respond to politics’ (2000, p. 4). He adopts a method that does this, but is primarily intended to govern an ideal. In this sense his method lies in the RU portion of the ideal/non-ideal continuum. Designing a conception of justice for an ideal allows him to ignore some practical difficulties and make simplifying counterfactual assertions (ibid., p. 73). However, he also warns that we should take careful notice of those simplifications, since he intends to apply his ideal theory into practice at a later stage. Indeed an attraction in favour of using a hypothetical market device is that it might be useful ‘in the design of actual political institutions’ (ibid., p. 72). In this sense a conception of justice must, for Dworkin, be ultimately implementable.
1.3.2 Interpretations of Feasibility Constraints

A purely intuitive ideal of infeasibility says something like ‘given our best understanding of human nature and its operation under different institutional arrangements, [a principle] is simply unrealizable’ (Valentini, 2011, p. 39). There are a number of factors that could influence whether a principle of justice is realizable in this way. Whichever conception a particular RU theorist adopts informs the methodological and justificatory principles compatible with her approach. So far I have only discussed those empirical factors that I take to inform the least controversial examples of infeasibility, sensitivity to which is ascribable to all existing RU theorists: those that describe the laws of nature, the circumstances of justice, or make successful completion extremely unlikely. There is broad agreement over this uncontroversial account, but discord over anything more substantial. Such discord isn’t an inherent problem since Realistic Utopianism refers to a range of approaches on the ideal/non-ideal continuum and so there is scope for disagreement over particular methodological principles.

There are, however, some instances where a thicker conception of feasibility or some other methodological adaptation is necessary for a theory to produce results that match our intuitions, as examples in the following sections show. Conceptual analysis is not, however, the aim of this section. I consider three types of empirical factors (contextual, psychological, and political). Each is selected because it yields intuitive counter-examples to the uncontroversial conception of feasibility and because (some) existing theorists in the literature have taken them to be relevant. I start by considering their intuitive plausibility and then discuss their treatment by some existing theorists. Although I conclude that substantive RU principles ought to be sensitive to these empirical factors, their sensitivity is not attributable to a concern with feasibility. Instead, each is a requirement of producing a conception of justice that is ultimately implementable for reasons other than infeasibility.20 These reasons are a commitment to contextualism, stability, and legitimacy. If correct,

---

20 This is not incompatible with theorists also adopting a thicker conception of feasibility that includes these kinds of factors. In such cases, theorists would have two reasons to give the factors a role in when developing a conception of justice.
each of these concerns imposes a methodological requirement onto RU theory in addition to the demand that conceptions of justice be feasible.

These requirements are more stringent than the uncontroversial account of feasibility because they impose further constraints on the types of methodological principles compatible with the RU approach. In some cases, existing theorists have inadequately dealt with these concerns. In others, the commitments are simply under-theorised so that it’s unclear how proponents of the RU approach should treat the different kinds of factors, and why the requirement to do so applies to them in particular. The insights I unveil here pave the discussion of the following chapters and the remainder of the thesis. They reveal the substance of what I take to be necessary in a fully-worked-out definition of the RU approach that follows from the core commitment as I define it.

1.3.2.1 Contextual Factors

A fact is contextual when it describes a non-universal set of shared circumstances. These circumstances might be shared amongst small communities, members of particular organisations (such as schools or churches), nations, or any number of other social forms. There are numerous kinds of facts that could describe these circumstances. Facts that describe natural resources, climate, economic situation, conventional practices, history, or infrastructure are all possible examples. Contextual facts’ non-universality distinguishes them from non-contextual facts; they do not apply to all humans in the world as we know it. This makes their effect on individuals contingent on the shared circumstances those individuals happen to be in. Contextual facts may or may not be permanent features of the shared circumstances.

In some cases it might seem that contextual facts can block requirements of justice. Imagine a context in which there is a large population and an inhospitable climate for growing crops. A plausible principle of justice is one that requires the state to ensure that everyone’s basic needs, including nutrition, are met. Whether or not this principle features in a conception of justice depends on the idealised society it is intended to govern. Let’s say that given some description of the features of the idealised society, we conclude
that it’s a plausible requirement. Ensuring basic needs are met is the sort of thing that governments should do, and it’s feasible given the (universal) circumstances of justice and probability. However, if we add information about the particular contextual circumstances of the state, our intuition might be that although it’s plausible, fulfilling this requirement can’t be a requirement of justice. The inhospitable environment apparently means that it’s infeasible for the state to fulfil the requirement to feed its citizens. Our intuitions about whether malnutrition constitutes injustice seem to depend on context. If there is genuinely nothing that these authorities can do to fulfil this requirement, then they are not acting unjustly.

This requirement isn’t, however, ruled out by the uncontroversial account of feasibility. This account takes only universal features of our world that define the laws of physics, circumstances of justice, and likelihood of success to be the appropriate context for feasibility assessments. The contextual facts that inform our judgement in the previous case are not included in this list. Strictly speaking, feeding everyone is possible, since there is enough food in the world and there are means of transporting it. As such, it is not impossible in the sense adopted by the uncontroversial conception of feasibility and there is a tension between our intuitions and the uncontroversial account of feasibility. If this conception delivers the wrong result in this case then theorists must either reconsider their conception of feasibility or adopt an alternative justificatory principle into their methodology that accounts for cases like this.

RU theorists that exclude contextual facts from their deliberations in this way are *universalist*. On their view, theorists may produce ultimately implementable conceptions of justice irrespective of the particular contextual circumstances of the society to which they are to apply. Their ‘aim is to discover principles of justice that can and should guide our judgement and our behaviour in all circumstances—principles that are universally applicable in the sense that whenever we have to assess the justice of a decision, or an institution, or a policy, we appeal, at base, to the same considerations’ (Miller, 2002, p. 7). Other theorists may accept that contextual facts have a methodological role. Some proponents of the practice dependency view think that ‘abstract
principles such as this [the basic liberal principle of the equal moral worth of all human beings] do not qualify as principles of justice, proper, but merely constitute general moral requirements that every conception of justice has to respect. The task of a suitable conception of justice is then to provide an account of what it means, for a specific practice, to respect such requirements’ (Banai, Ronzoni, & Schemmel, 2011, p. 50). Specific practices are contextual facts and have a role in formulating fundamental principles of justice on this view. These different views regarding contextualism represent different points on the RU portion of the ideal/non-ideal spectrum.

In chapter two I argue that RU principles ought to be sensitive to some contextual facts because their core commitment requires them to reject the universalist claim that contextual facts cannot contribute to a principle’s fundamental grounds. The reasoning in favour of context-sensitivity matches the justification for including feasibility considerations into substantive principles’ formulation. In this chapter, we saw that feasibility is a necessary condition of a conception’s being ultimately implementable. In chapter two I establish context-sensitivity as a second necessary condition. Although both requirements of RU theory share a justification, it’s not the case that context-sensitivity is a necessary requirement of feasibility. Rather, contextual considerations are necessary to ensure that a principle is sufficiently determinate for theorists to claim that their conception of justice is ultimately implementable. A principle is sufficiently determinate when it delivers a definite verdict regarding the justness of a particular action or institution.

This is a more plausible explanation of the intuition described above. Contextual facts are relevant for the RU approach to justice because they are necessary to make the results of deliberations determinate, not because they impose feasibility constraints. Our intuition that a conception of justice should nevertheless respond to such facts is explained by the level of determinacy required to make claims about ultimate implementability. This claim doesn’t preclude the possibility that RU theorists supplement their account of feasibility to include sensitivity to contextual facts. Doing so is not, however, a requirement of the RU approach. Sensitivity to contextual facts for
determinacy follows from the core commitment of the approach and accounts for intuitions similar to the case described above.

1.3.2.2 Psychological Factors

There are numerous psychological facts about humans that theorists may think can make the implementation of a conception of justice infeasible. Facts of human nature include things like motivational constraints, partiality, and epistemic limitations. These considerations can make an action impossible or extremely difficult without there being any physical constraints that violate the uncontroversial account of feasibility. Limited altruism is a psychological factor that is taken into consideration as one of the circumstances of justice and thus accepted as a constraint by many existing theorists and featured in the uncontroversial account. Other psychological factors that always have and always will apply to all humans in any possible world can also impose constraints according to the uncontroversial account. The purpose of this subsection is to consider whether other, or perhaps all, psychological constraints similarly restrict the content of justice for feasibility reasons or otherwise.

This intuition about the relevance of psychological characteristics of humans is widely shared by existing theorists. Sam Freeman describes the way that Rawls and Scanlon observe psychological factors as a constraint; both of their methodologies are intended to apply to persons ‘with natural human tendencies and who are subject to normal conditions of social life’ (2009, p. 51). Similarly, Dworkin’s equality of resources (2000, p. 108) makes distributions ‘ambition-sensitive’ thus making individuals’ motivations, aims, and preferences relevant to a just distribution. Substantive principles that ignore the fact that people have distinct ambitions will fail the envy test and as such are not acceptable principles of justice. Nagel, too, highlights this aspect of feasibility, saying that ‘if real people find it psychologically very difficult or even impossible to live as the theory requires, or to adopt the relevant institutions, then that should carry some weight against the ideal’ (1991, p. 21). Carens says ‘you cannot use standards that no one ever meets or even approximates in their actual behaviour…moral norms should not stray too far
from what most actors are willing to do much of the time’ (1996, p. 158).

Freeman describes the intuition similarly:

A sincere and conscientious utilitarian is a person who should have no special concern for himself or herself and who impartially promotes everyone’s interests. But it is beyond our capacities to forgo whatever aims and interests we have and develop a settled disposition always to take everyone else’s interests equally into account and act to maximize overall utility. To contend that we nonetheless ought to do so when we by nature cannot is to make an unreasonable demand. How can a reasonable morality demand something realistically impossible for people? (2009, p. 47)

The consensus is that although there is no physical obstacle that prevents utilitarian calculations and other psychologically difficult things from informing our actions and judgements, a conception of justice that requires us to act in this way is in some sense undesirable. The uncontroversial conception of the RU approach cannot account for all instances of this undesirability since it does not render all such actions to be infeasible. Impartiality towards one’s interests in some cases is not impossible, does not violate the circumstances of justice, and is not unfeasibly improbable (given that the conception of probability is conditional upon trying). Despite this, a conception of justice that is ultimately implementable into the real world must be in some way sensitive to these intuitions in cases where they should clearly block requirements (for instance impartiality between saving your life and that of a stranger). Some might reject this claim on the basis that other actions that imply partiality, for example parting with one’s money, might be psychologically difficult but can still plausibly be requirements of justice. So it seems that some, but not all, aspects of human psychology can shape the content of RU justice.

Note that these examples are distinct from the lazy father (who did not want to go and see his daughter’s hockey game) because they are supposed to reflect traits that it’s plausible to say we all have; they are characteristic of humans.
Despite widespread acceptance of the idea, it’s unclear whether these theorists take sensitivity to psychological facts to impose a feasibility constraint or to shape the content of principles of justice in some other way. As such, it’s worth considering the methodological implications of such a commitment in more detail.

This intuition about psychological facts acting as constraints appears to conflict with RU theorists’ permission to idealise their subjects. This complicates the methodological implications further because it means that a particular limitation on idealisations must be established.\textsuperscript{22} It is unclear which features of humans theorists may idealise whilst remaining loyal to the aim of producing ultimately implementable substantive principles. The fact that not all human attributes block requirements suggests that an appeal to infeasibility does not resolve this problem. In terms of feasibility, acting according to utilitarian principles might impose the same psychological demand on me as giving away money. Yet one is rejectable on this basis and the other is not. Theorists must, then, have specific arguments that show why some psychological facts shape the content of conceptions of justice, while others do not.

In chapter three I consider further the difficulty of deciding which psychological features should and shouldn’t constrain conceptions of justice and suggest that the minimum requirement for RU theorists in this regard is that they demonstrate their conceptions of justice to be \textit{stable}. Stability requires a conception of justice to be able to order society over time and so theorists must consider the motivational capacities of citizens to do that. Moreover, RU theorists should ensure that their conceptions are stable for the right reasons. This means that they must show them to be compatible with idealised citizens motivated by a sense of justice, a common idealisation of human motivation. The stability requirement is necessary for the simple reason that it’s implausible

\textsuperscript{22} Accepting some psychological factors as constraints does not require RU theorists to consider a real world constituency as is the case with NI theory. This is because RU theorists will still be considering a particular conception of humans. In the real world there will always be people who do not fit a particular conception, even with something as uncontroversial as limited altruism, and so RU theorists idealise. This idealisation distinguishes their approach from the NI alternative.
for RU theorists to reject it given their commitment to feasibility. I argue that stability is best understood as a stronger and separate consideration to feasibility, but the reasoning in favour of adopting them both is the same. At the same time, the stability requirement allows RU methodologies to deliver intuitively appealing results—that account for the intuition discussed at the beginning of this section—when it comes to psychological facts, and sets a principled limit on the use of idealisations that matches the essential nature of the approach established in the thesis thus far.

1.3.2.3 Institutional Factors

‘Institutional factors’ may cover a number of empirical facts that some theorists might think influence a conception’s feasibility. By ‘institutional factors’ I mean any empirical circumstances that describe global or national political, social, or economic institutional arrangements today. Such factors are excluded from the uncontroversial conception of feasibility because the circumstances they describe are ordinarily thought to be temporary. The institutional features of our societies and the world have changed many times and are likely to continue to do so. As such, they don’t affect whether a conception of justice is ultimately implementable when judged on the uncontroversial conception of feasibility alone.

Hamlin and Stemplowska call a restrictive account of feasibility that responds to all the factors that constrain choice here and now as a ‘short-run’ feasibility frontier (2012, p. 56). Although this idea doesn’t resemble the intention of RU theorists, it might be plausible in the following sense. Recall that feasibility is important to RU theorists at all because their core commitment is to develop conceptions of justice that are ultimately implementable. So far, I have not considered the idea that this consideration requires some acknowledgement of where and when we are starting. Rather than asking themselves ‘could this be implemented’, perhaps RU theorists should, instead, ask ‘could we implement this’. The latter question makes the relevance of institutional arrangements

---

23 Institutional facts may be contextual or universal. The democratic system in the UK is an example of the former and the existence of national borders is an example of the latter. In this sub-section, I am interested in their institutional, rather than their potentially contextual, nature.
clear and the idea that they might inform feasibility more plausible. Carens describes the intuition that these concerns are relevant to feasibility when he says we ‘should accept as a starting point the division of the world into states that are, at least formally, sovereign and independent…on the grounds that these sorts of proposals [those that require radical changes to the institutions] have no chance of being implemented or even being given serious consideration’ (1996, p. 158). The most significant difference between these two considerations is that acknowledging the starting point requires theorists to consider the moral and non-moral costs associated with implementation. The uncontroversial conception of feasibility does not require theorists to consider such costs. The following example helps to clarify the different considerations at stake here.

Consider whether travelling to London from Manchester via rail in 68mins is feasible even though it’s currently impossible. The fact that it is currently impossible doesn’t, on the uncontroversial conception, mean that it is infeasible. It could, after all, be possible in a few years if the right infrastructures and supporting institutions are built. However, the only reason that we know this is because the planning has already been done. Oddly, if there weren’t current plans for a high-speed railway line, then our intuitions regarding its feasibility might change (even though the possibility of those plans coming to fruition is the same regardless of whether they have, in fact, been planned). Consider whether you think it is feasible to travel by rail from London to Rotterdam (roughly the same distance as Manchester) in the same time. Intuitively, this seems less feasible, but why? Presumably it’s possible in the sense relevant for the uncontroversial conception. Perhaps it is still feasible, but only if time is not a constraint. On the other hand, it may be the case that the time it takes to make an action possible isn’t the relevant variable to our judgement, but rather its costs. Costs can be understood as either moral (i.e. damaging the environment and killing animals) or non-moral (i.e.

---

24 If the presence of the North Sea is responsible for the difference in your intuitions, then ask instead whether it’s feasible to take a train to Rotterdam in less than three hours (Paris is also 200 miles from London and this journey takes 2h30min). The example is supposed to reflect the idea that the existence of a plan to change infrastructure somehow changes our intuitions, which is peculiar.
financial). Perhaps it would be possible for me to make the rail journey if everyone in the UK stopped what they were doing in order to work on the new high-speed line. However we think each of these considerations should bear on this particular case, what is at least clear is that institutional concerns invite a number of complicated and overlapping considerations. Clearly, admitting these considerations into a conception of feasibility would drastically change the content from the uncontroversial account.

Rääkkä explicitly rejects institutional feasibility constraints when he says ‘a theory of justice suggests social arrangements that are feasible, it does not follow that they are politically [institutionally] feasible, too’ (1998, p. 28). One reason that he gives for this is that for a decision to be politically feasible, it has to be true that it can be acted upon reasonably quickly. The same is not true of the feasibility condition in political theory. For him, the two concepts are distinct and our intuitions about whether something is implementable in the short-term do not inform our intuitions about whether a proposal is feasible (in the political theory sense). We shouldn’t understand sensitivity to current political, social, and economic institutions to be a necessary feature of RU conceptions of justice on feasibility grounds, at least.

Not only does this institutional conception of feasibility fail to reflect the way existing theorists like Rääkkä use the concept, but also it’s simply implausible. It would be very difficult to establish a threshold whereby a principle that’s implementable in five years is feasible, but not if it takes, say, ten. We would also have to appeal to a calculation between feasibility and desirability to ensure that both moral and non-moral costs could be considered in implementation assessments. In particular, an institutional account of feasibility is implausible for the RU approach given the aspirational nature of the theory. Making it sensitive to feasibility constraints like these would make the theory subject to the same criticisms as the NI approach (that arbitrarily privilege the status quo). The RU approach is intended to guide the constituents of a utopia and it’s unclear that sensitivity to institutional arrangements is necessary to this end. Even if we consider more established institutional facts like the presence
of an authoritative state in every country of the world today, it’s unclear that this is sufficient to show state-less conceptions of justice to be infeasible.

The purpose of this discussion is to show that RU theorists needn’t allow institutional concerns to shape their conceptions of justice for feasibility reasons. This doesn’t mean that RU theorists cannot observe any political constraints, merely that their doing so is not attributable to a feasibility concern. In chapter four I consider the facts of the permanent and coercive nature of political power. I say that because of these facts, a necessary condition of ultimate implementability is that a conception is legitimately implementable. The coercive nature of implementation is a permanent fact of the nature of political life. Coercion is presumptively wrongful and thus incompatible with implementing a conception of justice unless doing so can be shown to be legitimate. RU theorists must show their conceptions of justice to be legitimately implementable so that its implementation is not wrongful and not because they are otherwise infeasible.

Within the discussions of different factors that feasibility might require theorists to consider, I’ve earmarked three ways that RU theorists must modify their approach. These modifications don’t commit them to a more substantial conception of feasibility, but rather to alternative commitments (to contextualism, stability, and legitimacy) that follow from the core commitment to developing ultimately implementable conceptions of justice. I discuss each of these in the remaining chapters of the thesis. They either i) constrain the types of justificatory and methodological principles permitted by the theory or ii) demonstrate a particular justificatory, methodological, or substantive principle that must feature in any RU theory of justice.

**Conclusion**

The RU approach to justice represents one of several that may be distinguished by their commitment to a particular function and aim of a theory of justice. For proponents of this view, a theory’s function is to order social arrangements and provide action-guiding recommendations. Other approaches on the
ideal/non-ideal continuum share this view, but interpret the requirement to produce action-guiding recommendations differently. For RU theorists, the aim is to develop substantive principles for a moderately idealised world, rather than the world here and now, or an infeasible utopia. All pursuits on the continuum are worthwhile and plausible interpretations of how a conception of justice can order social arrangements, but RU is of particular interest. This interest stems from the prevalence of the approach in the literature, and the intuition that knowing how just we can possibly be is of some intrinsic value. For these reasons, the remainder of the thesis is focused on expanding its working definition to reveal a more determinate range of approaches.

The way that RU theorists interpret their aim is what I call the ‘core commitment’ of the approach. This commitment establishes the broad methodology that existing theorists share. Theorists use idealisations to develop substantive principles that govern a moderately utopian society, but limit them by acknowledging and responding to feasibility constraints. Feasibility constraints are an important feature of their considerations because their aim is to develop conceptions of justice that are ultimately implementable. Existing theorists agree only over an uncontroversial conception of feasibility.

In the final section, I introduced three kinds of empirical concerns that theorists might include in a more substantive conception of feasibility. Doing this allowed me to demonstrate briefly the extent of disagreement over these areas and the way in which existing theorists have differed with respect to them. The scope of disagreement over feasibility constraints and idealising assumptions means that there is no determinate RU method to develop substantive principles of justice. This, in itself, isn’t a problem so long as any formulation within the range remain loyal to the aim of RU conceptions of justice. In the chapters that follow, I argue that there are much more determinate and stringent requirements of the RU approach that follow from the core commitment. In particular there are three concerns that shape the method; contextualism, stability, and legitimacy. I turn to contextualism first.
2. Contextualism

Introduction

In the previous chapter we saw that RU theorists must acknowledge feasibility constraints when developing a conception of justice. I described an uncontroversial account of feasibility to which all RU theorists are committed. According to this account, an action’s infeasibility in a particular context is insufficient to say that it is infeasible in the sense that informs deliberations over RU principles. Such actions can therefore be requirements of justice.

In this chapter I claim that some fundamental principles of an RU theory of justice should be context-sensitive even though contextual facts do not undermine their feasibility. I call this position *contextualism*. Fundamental principles are the non-derivative substantive principles that form a conception of justice. Their context-sensitivity is not attributable to infeasibility, but rather the result of contextual facts forming a part of their grounds. Contextual facts must form part of the grounds for some fundamental principles because they ensure that the conception’s requirements are sufficiently determinate. Determinacy is necessary for fundamental principles so that theorists can show them to be ultimately implementable, as per the core aim of the RU approach. In this way the reasoning in favour of feasibility matches the reasoning in favour of contextualism; they are both necessary conditions of producing an ultimately implementable conception of justice. I defend *weak contextualism* which is the idea that context may partly ground some substantive principles of justice rather than a stronger claim that makes contextual facts the grounds of
some or all principles entirely. My arguments do not rely on a particular account of the kind of contextual facts that form principle’s grounds.

Accepting the contextualism requirement affects the moderately idealised society to which RU conceptions of justice are to apply. It must be described in a way that includes some contextual facts as well as those that inform the uncontroversial conception of feasibility. The justification of fundamental principles then draws on these features of the society. In this way, contextualism restricts the range of justificatory principles compatible with the RU approach because it requires theorists to reject those that rule out the possibility of contextual fact’s contribution to justifying substantive principles of justice.

The structure of the chapter is as follows. In section one, I say more about the meaning of contextualism and contextual facts, and the nature of the contextualist requirement I support. Then, in section two, I argue that RU theorists must reject the universalist claim that contextual facts cannot contribute to the grounds of fundamental principles of justice. I then refine my preferred view in section three and demonstrate the implausibility of stronger variants. With the defence of my weak contextualist claim complete, I demonstrate examples of weak contextualism in existing literature in section four. Finally, in section five, I tentatively introduce the suggestion that people’s views about justice ought to be a contextual fact that RU theorists include in their account because doing so is a requirement of a substantive commitment to treating people equally.

2.1 Contextualism

Contextualism in political theory is the view that fundamental principles of justice (i.e. a conception of justice) ought to be in some way sensitive to contextual facts. Fundamental principles can be sensitive to contextual facts if

---

25 I use the term ‘fundamental’ to refer to the subset of substantive principles that are non-derivative. The distinctive claim of contextualism refers to these fundamental principles, although I sometimes refer to substantive principles, or conceptions of justice more generally.
their content, grounds, or scope depends on, or is in some way constrained by, contextual facts. Contextualism contrasts with universalist claims that deny contextual facts can contribute to fundamental principles in these ways.

Contextual facts describe non-universal shared circumstances. That is, they do not describe the circumstances faced by all humans in the world as we know it; whether or not a contextual fact affects a particular individual is a matter of happenstance. Non-universality is a necessary, but not sufficient condition of a contextual fact. A second necessary condition is that contextual facts describe circumstances shared with people in your proximity. Circumstances may be shared amongst communities, members of particular organisations (such as schools or churches), nations, or any number of other social forms. I will often have in-mind the idea that a nation-state represents the limits of a shared circumstance, although I do not claim that these are the only type of context that political theorists might plausibly be interested in.

There are numerous kinds of facts that could describe the shared circumstances of a particular nation-state. Facts about the economic situation, availability of natural resources, existing institutions, cultural beliefs, infrastructure, and history are all plausible examples of facts that contextualists might argue theorists should make their principles sensitive to. So, a non-universal fact such as men being unable to give birth is not a contextual fact—it’s not a circumstance shared with those around you—but the fact that Haitians are at risk from extreme weather, that honour killings are a common cultural practice in many regions of Pakistan, and that the UK currently owes a large financial deficit are all contextual facts. Each of these examples represents a different kind of contextual fact from the list above.

To say that all of these are examples of contextual facts is not to say that they should all feature in the reasoning and approach of a contextualist theorist. I postpone the question of which kinds of contextual facts principles of justice should be made sensitive to until the end of the chapter. My argument in favour of the claim that RU theories should be contextualist does not depend on theorists adopting a particular kind of contextual fact; I merely claim that substantive principles ought to be sensitive to some. Arguments about which
kinds of contextual facts are the most appropriate for RU theorists (for example the claim that geographic facts must feature in their reasoning but that existing cultural practices needn’t) can be made separately to the claim that sensitivity to some is important. In this chapter I argue that RU theorists have reasons beyond feasibility concerns to adopt contextualist grounds. I say more about what it means for a principle to be sensitive to a contextual fact and what happens, exactly, when a theorist includes contextual facts into her approach, before showing why RU theories ought to be contextualist in the aforementioned way.

I have said already that fundamental principles can be sensitive to contextual facts if their content, grounds, or scope depend on, or are in some way constrained by, contextual facts. So, one way that a theory of justice is contextualist is if contextual facts form part of the justification for some fundamental principles. In this case the fundamental principles are contextually-grounded. The following principle is an example of a contextually-grounded substantive principle. ‘Distribute education in society X according to the views of X’s members’. A principle is contextually-grounded when it only generates recommendations with some input from contextual facts (in this case the views of members of X). In this case the context determines the content of the principle, but the justification of the principle needn’t refer to contextual facts. Contextual facts could also inform or constrain a contextually-grounded substantive principle, rather than fully determine it. The scope of a substantive principle (for example ‘distribute according to need’) is context-sensitive if and only if its application varies according to context. For example, the context of a modern liberal democracy could mean that the scope of the need principle extends to education (i.e. need is the basis of our distribution of education). In a different context from a modern liberal democracy, for example one in which education is scarce, we might think that a different distributive criterion (i.e. maximising utility) is more appropriate to distribute education. This narrows the scope of the need principle since it no-longer includes education.

My contextualist claim is specifically regarding the grounds of justice. The grounds of justice are reasons that support particular substantive claims. They
are specified by the justificatory principles in a theory of justice. For example, the justificatory principle ‘true principles are those that best account for our most firmly held considered convictions’ identifies our most firmly held considered convictions as the grounds of justice. The grounds of the distribution of education or supporting the need principle in the previous examples were unspecified and may or may not refer to contextual facts. In this chapter, I argue that RU theories should contain a contextualist-in-content justificatory principle—one that makes the justification of fundamental principles sensitive to their context so that their context forms (at least a part of) the principles’ grounds. The justificatory principles need not be contextually-grounded themselves. ‘True principles are those that best account for the considered convictions of those to whom they are to apply’ is an example of a contextualist-in-content justificatory principle. In this case, the contextual facts surrounding peoples’ considered convictions are the reason that theorists adopt a particular substantive principle over alternatives.

Focusing on the grounds of justice reflects a trend in the literature that identifies the disagreement between contextualists and non-contextualists to be methodological rather than substantive (Valentini, 2011, p. 399). Contextualists don’t necessarily disagree with non-contextualists over what the correct substantive principles of justice are, but they do disagree over their grounds. Non-contextualists and contextualists might agree for example, that in the UK, justice requires equality of opportunity, but the reasons in support of that claim would differ.26 I am interested in this methodological disagreement because the purpose of the thesis is to establish the methodological commitments of the RU approach. To argue that the fundamental principles of a theory are contextualist-in-content would be to make substantive claims about the nature of justice, which is outside the scope of the thesis.27 My position is to argue

26 The non-contextualist view would give reasons in favour of equality of opportunity independent of the context of their application, whereas the contextualist would endorse the principle on the basis of support from some contextual facts about the UK.
27 There may be some relation between grounds, content, and scope, such that contextualism in terms of the former entails contextualism in terms of the latter in some or all cases. I leave this question aside and only note here that I take the claims in this chapter to apply regardless of such a relation.
against the particular universalist methodological claim that contextual facts cannot ground substantive principles of justice, rather than a substantive position that says a principle of equality, for instance, is universally true. As such, I am defending the particular claim that justificatory principles should be contextualist-in-content, and therefore that fundamental substantive principles should be contextually-grounded.

The most common form of contextualism from recent literature is practice dependence. This view is a good example of a methodology in which fundamental substantive principles have contextual grounds. Proponents of this view think that conceptions of justice ‘are constructed by interpreting the point of the specific practice they aim to regulate’ (ibid., p. 400). Practices are empirical facts that describe the non-universal shared circumstances of the people within societies and as such meet the criterion of contextual facts. Examples of practices include the functional role of political society which ‘is some form of cooperation between citizens’ and the functional role of international society which is ‘some form of cooperation between states’ (Meckled-Garcia, 2013, p. 97). Meckled-Garcia acknowledges the contextual nature of these facts when he says that they are ‘contingent and historically specific’ (ibid., p. 103). Although many of the practices these theorists are interested in are commonplace in today’s world (for example, the centralised legal system of nation states), they are contingent because their functional role is interpreted through the point of view of the participants (Sangivanni, 2008, p. 148). This can deliver different results depending on participants’ context. As such, they fulfil the necessary and sufficient conditions of a contextual fact. Sangiovanni, an important proponent of the practice dependence view, describes fundamental substantive principles as ‘depending’ on practices. This means that ‘first principles of justice…can vary according to the institutions and practices which they are meant to regulate’ (Sangivanni, 2008, p. 138). This relation describes the way in which practice dependence theorists ground their principles on contextual facts. On their view, a practice must be justly regulated, but what makes its regulation just partly depends on its point and
purpose. As such, the point and purpose of practices feature in the grounds of the principles that govern them. The practice dependence approach is compatible with the contextualist claim I defend in this chapter, although the claim is not so narrow as to specify this approach as the only contextualist option for RU theorists.

I defend a particular idea of contextualism and the role of contextual facts in the grounds of justice. I call this view weak contextualism. According to weak contextualism, context forms a part of the grounds for some fundamental principles of justice. I don’t specify a particular justificatory claim (a particular standard that specifies what makes a principle just or unjust). Instead, my claim imposes constraints on the justificatory principles compatible with an RU methodology. There are two senses in which this contextualism is weak, each of which broadens the range of compatible approaches. First, it does not demand that context forms the grounds of every substantive principle in a conception. This means that principles that prohibit slavery, for example, could have entirely non-contextual grounds and feature in the conception. Second, it acknowledges that non-contextual grounds justify principles alongside their contextual justification. For example, substantive principles that specify voting systems might justifiably vary according to context and also respond to their non-contextual grounds (by, for example, protecting the right to an equal say).

Weak contextualism is compatible with some versions of practice dependency and other existing contextualist methodologies, as I demonstrate in section four. The most significant constraint this requirement imposes is to reject the universalist view that contextual facts cannot contribute to a fundamental principle’s grounds. I defend this view in the following section.

2.2 Contextualism Over Universalism

Contextualism is the view that the grounds of some fundamental principles of justice are given by the context to which they are to apply. This particular claim

---

28 Since these theorists interpret practices through the perspective of participants, they will always have a contextual nature. This is compatible with my view but it is also worth noting that (non-universal) practices could also be contextual facts without this interpretive element.
is distinctive to contextualism; all theories of justice might generate some contextually-grounded principles, but they may be derivative rather than fundamental. Contextualist literature reflects this distinctive element; ‘contextualists claim that there is a relationship of appropriateness between context and principle that is primitive in the sense that it cannot be explained by appeal to some more fundamental principle that applies universally’ (Miller, 2002, p. 11). Defining contextualism in this way does not entail that we reject universalism entirely, but rather the following particular claim that I attribute to a view called strong universalism.

According to strong universalism, contextual facts cannot contribute to a fundamental principle’s grounds. Weak universalism, on the other hand, could admit both contextual and non-contextual grounds into their approach. The distinction between weak and strong universalism is methodological rather than a substantive claim about the scope of conceptions of justice. Numerous familiar conceptions of justice are compatible with strong universalism since they don’t appeal to any contextual grounds. For example: conceptions grounded in the basic intuition that we should mitigate the effects of bad brute luck; libertarian accounts that assign basic self-ownership; and various forms of classical utilitarianism (Sangiovanni, 2008, p. 139). For these theorists the ‘aim is to discover principles of justice that can and should guide our judgement and our behaviour in all circumstances—principles that are universally applicable in the sense that whenever we have to assess the justice of a decision, or an institution, or a policy, we appeal, at base, to the same considerations’ (Miller, 2002, p. 7).

These theorists object to contextual facts forming the grounds of fundamental principles of justice arguing that ‘to assign any greater role to institutions and practices—to allow them, as I have said, to influence the formulation and justification of first principles of justice—is a fundamental mistake: constraining the content of justice by whatever social and political arrangement we happen to share gives undue normative weight to what is, at best, merely the product of arbitrary historical contingency or, at worst, the result of past injustice itself’ (Sangiovanni, 2008, p. 137). This kind of objection is similar to
Meckled-Garcia’s view regarding the practice dependency approach. He says: ‘a feature of the practice is a relevant consideration as to what principles are relevant to the practice, but that is itself for a moral reason: a reason explaining why that feature is important and triggers moral considerations. The principles themselves are not formulated or constrained by reflection on socially accepted functional roles’ (Meckled-Garcia, 2013, p. 117). The universalists’ view is that fundamental principles are justified entirely by non-contextual considerations. Their view appears to be compatible with the broad methodological requirements of the RU approach established so far.

In this section, I defend the claim that RU theorists must reject strong universalism. They must admit contextual facts into the grounds of some of their fundamental principles when those principles would otherwise be indeterminate. Determinacy of fundamental principles is important for RU conceptions of justice because it is necessary for them to show their principles to be ultimately implementable. RU theorists need to take account of these contextual facts because they’re essential to the nature of the RU approach, not because they can render a conception of justice to be infeasible (note that this doesn’t preclude the possibility where there are some cases in which contextual facts impose feasibility constraints). In section three, I argue that only weak contextualism yields plausible justificatory principles and that stronger varieties fail.

2.2.1 RU Theory and Determinacy

As we have seen, a fundamental characteristic of substantive RU principles is that they be *ultimately implementable*. This means that it must be possible for these principles to eventually guide societies in the world as we know it. This feature distinguishes RU theory from the Utopian alternative. The requirement for principles to be ultimately implementable grounded the claim that RU principles of justice must be sensitive to feasibility constraints. RU theorists are interested in pursuing principles that push the limits of practical possibility. Recalling this reason reminds us of the sense in which RU theorists are concerned with the world-as-we-know-it despite the fact that they are developing principles for a moderately idealised society.
The minimal conception of feasibility introduced in the previous chapter imposes constraints determined by non-contextual facts only. Such facts are the universal features of our world that define the laws of physics, circumstances of justice, and probability of success to be the appropriate context for feasibility assessments. I leave open the option for RU theorists to supplement this account with a thicker conception of feasibility that includes contextual facts as the source of possible constraints. I discussed this possibility in chapter one.

Although there are some plausible cases in which contextual facts impose feasibility constraints, this isn’t always the case, and as such, a thicker conception of feasibility is insufficient to show that RU theorists must reject the strong universalist claim and accept the contextualist requirement. When the content of substantive principles is indeterminate, it’s important for RU theorists to take context into account even though its impact on feasibility may be absent or negligible. In this sense, the reasoning in favour of contextualism is the same as that of feasibility—they are both necessary conditions of ultimate implementability—but meeting the requirements of one is insufficient to meet the requirements of the other.

Without determinacy in the requirements of a conception of justice, RU theorists cannot show their conceptions to be ultimately implementable. Indeterminacy is commonly accepted as a flaw of approaches to justice. Onora O’Neil cites this problem as a common criticism of Kant’s Categorical Imperative ‘which simply lacks determinate implications for action. Mill speaks for many when he alleges that Kant “when he begins to deduce from this precept any of the actual duties of morality...fails almost grotesquely...”’ The supposed failure of Kantian formalism is neatly summarized in Peguy's acid quip “Le kantisme a les mains pures, mats il n'a pas de mains” What is the point of 'hands' kept clean by grasping nothing?” (O'Neill, 1987, p. 57). I hold RU conceptions of justice accountable to the following standard of determinacy that I’ve adapted from a definition intended for moral theories more generally (Timmons, 2013, p. 13). To say that a conception of justice is
determinate is to say that its principles yield definite verdicts about the justness of particular institutional arrangements or citizen’s/state’s behaviour.29

Consider a fundamental principle that says justice requires us to ‘be respectful of persons’. This principle is indeterminate because it doesn’t tell us whether particular institutional arrangements or political behaviours are unjust. It’s unclear whether examples like imprisoning people or means-tested welfare are compatible with or prohibited by this principle.30 Proponents of the practice dependent approach share the view that this kind of indeterminacy is problematic. They hold the position that ‘abstract principles such as this [the basic liberal principle of the equal moral worth of all human beings] do not qualify as principles of justice, proper, but merely constitute general moral requirements that every conception of justice has to respect. The task of a suitable conception of justice is then to provide an account of what it means, for a specific practice, to respect such requirements’ (Banai, Ronzoni, & Schemmel, 2011, p. 50). So, before we accept the requirement to treat people with respect as a fundamental principle of justice, it must be expressed in a way that is sufficiently determinate. This is particularly important for RU theorists because their aim is to produce principles that are ultimately implementable. The determinate requirements of the principle must be known before RU theorists can show it to be ultimately implementable. This principle could require any number of particular principles. For instance, the imposition of the harm principle, the right to vote, a particular resource allocation, or limitations on the scope of state’s power.

29 Note that this requirement does not entail that every particular institutional arrangement or policy must be worked out for a conception of justice to be determinate. Rather, for a conception to deliver a verdict, it only has to be the case that it’s clear whether a particular institutional arrangement or policy is compatible with or would undermine the requirements of justice. Theorists/lawmakers can then choose between these compatible options.

30 We might consider some of the examples I discuss in this section to be more aptly described as derivative rather than fundamental principles. The relevant point in this section is that fundamental principles must be determinate. It is arguable whether the various examples I give of the ways in which a principle of respect could be determinate are fundamental or derivative. I assume, for now, that they could be expressions of the fundamental nature of the principle. I want to highlight here their conceptual possibility rather than defending a substantive account of respect.
As we have seen, feasibility considerations are built into the RU methodology. This means that RU theorists must reject any interpretations of the requirements of the respect principle that are infeasible. Requirements cannot violate the circumstances of justice or other features of feasibility. An allocation of resources incompatible with moderate scarcity would, for instance, be rejected. Although this narrows the possible set of requirements stemming from a principle of respect, it is insufficient to ensure that it yields determinate guidance. Many, if not all, of the determinate possibilities I listed in the previous paragraph are within the constraints of the feasible.

2.2.2 Determine Principles and Contextual Facts

The only way for an RU theorist to produce sufficiently determinate principles is for her to permit contextual facts to form part of principles’ grounds. Determinacy requires theorists to produce a conception of justice that yields definite verdicts about the justness of particular institutional arrangements or citizen’s/state’s behaviour. If the grounds of a conception of justice can be partially given by its context, then this requirement can be met. Contextual facts are significant in this way because the institutional arrangements or political sets of shared circumstances that we are familiar with—those that theorists use to flesh out the details of their idealised societies—are contextual. The RU theorist could, for example, describe the idealised society to which her principles apply as one that shares the institutions we have in modern, liberal democracies today. Doing so allows her to consider the implementation of the conception of justice and specify the particular requirements of, for example, the requirement to treat people with respect. Our institutional infrastructure’s compatibility with the idea that respect requires citizens have the right to vote draws support for that determinate interpretation of the ideal. Since the implementability of this conception features as a part of its justification, the contextual facts surrounding it form a part of its grounds, too.

Universally grounded conceptions of justice purport to be justifiable independently of context, but RU theorists can’t assess their plausibility because we don’t know what their guidance is or would be without the existence of at least some contextual facts. The kinds of universal facts that
theorists would have to make a judgement on the basis of are human fallibility, moderate scarcity of fundamental resources, and a 24-hour day. If a theorist took her conception of justice to govern a world characterised by universally true circumstances only, then it would be impossible for her to show it to be ultimately implementable into any version of the real world as we know it. This is because of the lack of substance of universal facts. Even an empirical fact about something as seemingly uncontroversial as life expectancy is not universal or permanent. In some contexts, life-expectancy is much higher than others and the fact that it has increased over time and increased the size of our ageing population in the UK raises real distributive questions that we may not have previously considered. If theorists only acknowledge a universal fact about life-expectancy (that the average is 67.2), then we might question the justifiability of RU conceptions of justice in contexts with a much higher/lower figure. This is not to say that RU conceptions of justice should apply beyond the moderately idealised society for which they are designed, but rather that theorists cannot develop any determinate conceptions of justice unless some contextual facts are included into the deliberative process.

Banai, Ronzoni, and Schemmel make this point with regards to practices as contextual facts. ‘Without drawing on any facts about basic practices that exist in the global arena, and hence without an eye to how, and to which agents, cosmopolitan [universal] principles of justice would be applied, ultimately leaves the principles themselves without adequate support’ (Banai, Ronzoni, & Schemmel, 2011, p. 54). Contextualist conceptions of justice, on the other hand, ‘have a much better chance than PIA [practice independent] views of coming up with practically relevant recommendations which can be sensibly evaluated, because they apply to (idealized versions of) existing practices about which we have by definition enough knowledge to carry out this evaluation’ (ibid., p. 55). What these authors call ‘sensible evaluation’ is particularly important to RU theorists because of their focus on the practical relevance of conceptions of justice (this focus is the result of their commitment to ultimately implementable principles).
Steven Wall points out that some contractualist theorists restrict the scope of their conceptions of justice to the context of modern liberal democracies. Doing so permits the associated contextual facts to form a part of their principles’ grounds. The reason that they do this is because for their method ‘to yield determinate results, the range of people to whom it applies must be limited’ (Wall, 1998, p. 4). This is because the methodology of appealing to reasonable rejectability relies on a set of shared understandings. If the conception’s application extends beyond the theorist’s context, then she cannot be sure of the content of such an understanding. Many familiar liberal approaches have this contextualist nature, although it often goes unacknowledged. Rawls’s reference to his account as the ‘special conception of justice’ reveals that he too shares the view that contextual facts are necessary to avoid indeterminacy (Rawls, 1999, pp. 54-5). According to Rawls, the content of the two fundamental principles of justice as fairness only become clear once we narrow down the context to a particular type of society, with sufficient resources to meet the demands of the first principle (ibid.). I talk more about this relatively unfamiliar aspect of Rawls later in the chapter.

Universalists might accept that a conception of justice has different implications for different contexts, but reject the idea that the contextual fact’s role in delivering this determinacy is justificatory. Similarly, practice-independent theorists reject the idea that the contingent, practice-mediated, relations in which we find ourselves should change or affect the justifying reasons and premises sustaining a particular conception of the content and scope of justice (Sangivanni, 2008, p. 139). On my view, contextual facts ground conceptions of justice by specifying the balance between the weight of normative requirements and contextual empirical considerations. In some cases contextual facts inform judgements about the plausibility of a requirement of justice, and in other cases normative requirements are independent of (and cannot be overridden by) conceptual considerations. Universalists reject the possibility of the former case. The existence of the latter case is no problem for RU theorists; in these cases the requirements of justice are independent of the context. If its requirements are always the same, then its guidance is
determinate and contextual facts needn’t feature in its grounds to show that it is ultimately implementable (although its implementation must still be feasible).

The following example demonstrates how contextual facts may contribute to the justification of a fundamental principle in this way. If a theorist develops principles within the constraints of universal facts only, then she might think that the claim that respect requires direct democracy could be successfully justified. She cannot, however, assess the plausibility of this ideal until she knows more facts about the contextual circumstances. Circumstances such as population size, citizens’ interests, and institutional infrastructure can all inform whether direct democracy is plausible. For example, the large size of most currently-existing populations might lead her to conclude that direct democracy is implausible (at least for the majority of policy decisions). Moreover, the difficult practicalities that come with a large population size could undermine the way in which adopting it purportedly treats citizens with respect. As such, the plausibility of direct democracy is unknown until we consider contextual facts and the content of a principle of respect is indeterminate.

Universal theorist’s rejection of this justificatory role for contextual facts is mistaken. Part of the reason theorists adopt or reject a principle of direct democracy over alternatives is support from the contextual facts. Without considering them, the principle was both indeterminate and unjustifiable for a non-universal setting (and, as such, inadmissible into an RU conception of justice). Since this role for contextual facts is essential to ensure that fundamental principles are ultimately implementable, RU theorists must reject strong universalism. However, as we saw, this does not mean that RU theorists are committed to the claim that fundamental principles can never be universally justified. We have not rejected the idea that some fundamental principles can have universal grounds whilst others have contextual grounds. In this section I have shown why RU theorists must admit the permissibility of contextual grounds for fundamental principles (in cases where such principles would otherwise be indeterminate). In the following section, I discuss particular variants of contextualism which describe both kinds of role more
clearly. I argue that only weak contextualism is a plausible approach to developing RU principles of justice.

2.3 Argument in Favour of Weak Contextualism

I have said that it must be possible for contextual facts to feature in the grounds of RU conceptions of justice in order for those conceptions to be sufficiently determinate. This leaves open the possibility of RU theorists adopting any contextualist-in-content justificatory principle. In this section, I argue that only weak contextualism yields justificatory principles suitable for RU theorists because stronger varieties fail to be plausible methodological approaches.

2.3.1 Weak Contextualism

Weak contextualism is the view that context forms a part of the grounds for some fundamental principles of justice. Weak contextualism is distinguishable from stronger alternatives via two variables. First, it requires only some, rather than all, substantive principles to be contextually-grounded. Second, it requires that context form only a part of, rather than all of, these principles’ grounds. The stronger forms of contextualism and my labels for them are summarised in the table below.

<table>
<thead>
<tr>
<th>Grounds Fully Given (contextual grounds are sufficient)</th>
<th>Grounds Partly Given</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All substantive principles</strong> (contextual grounds are necessary)</td>
<td>Full Contextualism</td>
</tr>
<tr>
<td><strong>Some substantive principles</strong></td>
<td>Limited Principle Contextualism</td>
</tr>
</tbody>
</table>

*Table 1 Forms of Contextualism*

When grounds are ‘fully given’ it means that contextual facts’ support is sufficient justification for a substantive principle. In this case the relevant contextual facts are the only thing a theorist must consider when deliberating over whether she should adopt a particular principle. If, for example, theorists
adopt a distributive principle of need into their theory, and the only reason that the need principle is selected over alternatives is because of a contextual fact (for example, the endorsement of those to whom the principle is to apply), then that principle has its grounds fully given by a contextual fact. When grounds are partly given, then other context-independent ideals (e.g. fairness) may partly justify the principle as well. According to weak contextualism, contextual facts are sometimes a necessary part of a principle’s grounds, but their presence is never sufficient. Contextual grounds are necessary for those substantive principles whose content is otherwise indeterminate. Stronger views either acknowledge that contextual grounds can sometimes be sufficient, or require them as a necessary condition of every principle’s grounds (or both).

An approach that instructs theorists to select all substantive principles on the basis of citizens’ views, and those views only, is an example of a fully contextualist method. If all substantive principles are partly justified by their coherence with people’s views and partly by a principle of fairness, then the method is an example of limited content contextualism. An approach that specified distributive principles to be justified by their coherence with people’s views but principles of citizenship by a universal commitment to equality, say, is an example of a limited principle contextualist method. An example of a weak contextualist method is any contractualist approach that allows reasoning to be constrained by both universal and contextual concerns. I discuss examples from the literature that are compatible with weak contextualism in the following section. In the following sub-sections I explain in more detail how each approach works using examples from the literature and show why theorists should reject them all but weak contextualism.

2.3.2 Full Contextualism

A full contextualist (FC) theory of justice requires the grounds of all fundamental principles of justice to be entirely given by the context to which they are to apply. Moral relativism is a fully contextualist view. Support or coherence with their context is necessary and sufficient grounds for substantive principles and non-contextual grounds do not exist.
One reason that we might reject FC is because contextual-grounds should not be a necessary condition of all fundamental principles of justice (substantive or otherwise). This is a problem because the methodological principles in a theory necessarily have non-contextual grounds. Any theory of justice that includes a contextualist-in-content justificatory claim will also include methodological principles that specify how to understand and interpret the relevant contextual facts. Such principles specify which contextual facts feature in substantive principle’s grounds and why. As James says (with regards to social practices as contextual facts) ‘the initial interpretive problem is this: among the many and various existing social practices that can plausibly be said to exist, which should we attend to?’ (James, 2005, p. 298). Theorists must know the answer to this methodological question before they can adopt any contextualist-in-content justificatory principle that specifies social practices to be the grounds of substantive principles of justice. Methodological principles are necessary, recall, because they tell theorists how to satisfy the justificatory requirements of a theory; they tell theorists what information about practices (or other contextual fact) they need. The content of methodological principles and the answer to James’s question depends on a substantive claim about why those social practices in particular are relevant. And the grounds of that substantive claim must be context-independent. Rawls, for example, adopts a methodological principle that specifies social practices prevalent in the basic structure of society to be the primary subject of justice (1999, p. 7). Such a principle includes the practice of representative democracy with the scope of principles of justice but excludes hierarchical structures within corporations or universities. Practices surrounding the basic structure are relevant, according to Rawls, because of their profound and persistent effects on citizens’ lives (ibid.). This reasoning is context-independent.

Moreover, any reasoning in favour of a methodological principle like this must be context-independent because it’s necessary for theorists to avoid an infinite regress and get their theories off the ground. A substantive claim about the relevance of particular contextual facts is necessary to specify the facts’ content

31 In the following sections we’ll see that there are further reasons.
and there is nothing contextual for this claim itself to be based upon; its grounds must be universal. Any theory that demands this kind of substantive claim to be contextually grounded will fail to get off the ground.

FC fails to be a plausible approach because every theory of justice has methodological principles that depend on substantive claims with non-contextual grounds. This objection does not rule out the methodological requirement that all substantive (rather than methodological) principles of justice must have contextual grounds. FC shares this feature with another approach on the matrix, Limited Content Contextualism, and so I return to the suggestion that contextual grounds are a necessary condition of substantive principles justification in the corresponding section. First, I consider a more plausible, although ultimately rejectable, alternative to FC. This alternative, Limited Principle Contextualism, avoids the objection I pressed against FC by embracing the idea that methodological principles require universal grounds.

2.3.3 Limited Principle Contextualism

Walzer’s complex equality (1983) is an example of a method that features both contextually-grounded and context-independent principles and thus avoids the (first) objection to FC. According to Walzer, the contextual facts that form substantive principles’ grounds are the shared understandings of those to whom the principles are to apply. He argues that the goods a conception of justice distributes cannot be idiosyncratically valued, but rather should be comprehended through the community in which they are recognised as relevant (Walzer, 1983, p. 7). As such, ‘all distributions are just or unjust relative to the social meanings of the goods at stake’ (ibid., p. 9). Other principles in the theory have non-contextual grounds. The fundamental principles that define the kind of membership(s) that produces shared understandings have the context-independent grounds of an ideal of moral equality (Gill, 1987, p. 37). This conception of membership, what Walzer calls the ideals of democratic citizenship, ensures that the understandings really are shared, in particular by all who are affected by them (ibid. p. 48). This substantive claim about the nature of relevant shared understandings informs the methodological principle that specifies ‘how we may determine when a
conviction of a number of community members is a shared understanding, and one authoritative enough to ground principles of justice’ (ibid. p. 51). And the contextualist-in-content justificatory principle instructs us to adopt it.

This outline describes the way that Walzer’s method is an example of what I call Limited Principle Contextualism (LPC). When a group of citizens conform to his ideal of democratic citizenship, the shared understandings of those citizens are the contextual facts that ground the substantive principles of justice that govern them. Shared understandings are necessary and sufficient to ground these substantive principles. This view is not fully contextualist because it permits the grounds of democratic citizenship to be context-independent.

Limited Principle Contextualism takes the following (general) form. Contextual facts are necessary and sufficient grounds for some substantive principles of justice (in Walzer’s case distributive principles grounded in the shared understandings of the good to be distributed). The grounds of other substantive principles may be entirely non-contextual (in Walzer’s case those that describe democratic citizenship). I call this approach limited principle contextualism (LPC) because only some principles within the theory are contextually-grounded. The justificatory claim of this approach conflicts with mine because it makes contextual facts necessary and sufficient grounds for some principles of justice. On my view, although contextual facts are a necessary feature of some principles’ grounds, their presence is never a sufficient condition of a justified substantive principle.

LPC is an implausible methodological approach because we ought to reject the claim that contextual grounds are sufficient to justify substantive claims of justice (this gives us further reason to reject FC). Contextual grounds are insufficient because they cannot guard against ‘objectionable principles’. Principles are objectionable when they conflict with our highest order context-independent moral beliefs (or, as Miller puts it, when they are beyond the limits of the ‘edge of our conceptual universe’ (2002, p. 23)). The following example from Miller’s *Two Ways to Think about Justice* helps to clarify this point. Let’s say that context is given by citizens’ shared understandings, as is the case in Walzer’s theory. Imagine that the citizens of a particular society adopt a
practice of awarding honours or other tokens of merit to people who are
judged worthy to receive them. The members of this society happen to value
Negritude, and as a result only permit those with black skin to receive awards.
Proponents of LPC (and FC) are committed to saying that the fact that
members of their society endorse this practice is sufficient grounds to adopt it
and for resultant inequalities to be justifiable. Walzer appears to agree with
such a view when he claims, for instance, that ‘argument simply is the appeal to
common meanings’ (Walzer, 1983, p. 29). For him, the only way to make
moral claims about the honours practice is to appeal to its social meaning. RU
(and other) theorists object because the criterion of honours endorsed by this
society conflicts with many versions of equal treatment and respect that form a
part of their deepest and most considered convictions. Such convictions are
universal in scope and transcend particular approaches to developing principles
of justice. Although inequalities that result from honours systems can be
justifiable when honours are deserved, there is ‘no imaginable sense in which
having black (or brown, or white) skin can make you deserving. We have
reached the limits of our conceptual universe.’ (Miller, 2002, pp. 22-3) As such,
we cannot endorse an approach to justice that yields principles that protect
objectionable practices like the one in this example.\footnote{We must reject the claim
that contextual facts can be sufficient grounds for any principle of justice since
their non-objectionability, which is universally grounded, also contributes.}

If we reconsider the universal justification of methodological principles
introduced in the previous section, then we can see a second way that universal
grounds inevitably contribute to substantive principles’ justification. I said in
the previous section that methodological principles necessarily have non-
contextual grounds and gave Rawls’s restriction of relevant social practices to
those in the basic structure as an example of such a methodological principle.
The point and purpose of these social practices cannot be the sole justification
for substantive principles developed on the basis of them. If they were, then

\footnote{It is beyond the scope of this thesis to establish a substantive account of objectionability that accounts for all kinds of considered conviction that may apply in this case. It is sufficed to say that examples like Miller’s describe some clear cases whose existence undermines the claim that contextual grounds provide sufficient justification for substantive claims of justice.}
the following claim could be true: ‘representative democracy is just because it’s a practice reflected in the basic structure’. This substantial claim is clearly contextually-grounded, but the justification in favour of basic structure practices having this status is universal. This universal justification necessarily features in the substantive principle’s grounds, too. So, the claim above could be reformulated as ‘…is just because it is a practice reflected in the basic structure and such practices have a persistent and pervasive effect on people’s lives’. As such, representative democracy’s grounds are both contextual and universal. If universally-grounded methodological principles always have this effect on the grounds of substantive principles, then Walzer’s Complex Equality is not, in fact, an example of LPC. His account of democratic citizenship contributes to the justification of substantive principles of justice, in addition to their contextual justification from shared understandings.\(^{33}\)

Finally, contextual-grounds cannot provide sufficient justification for the substantive principles of an RU theory, in particular, because doing so constrains recommendations to those found in the contextual status quo. This prevents the resulting conception of justice from producing the aspirational results that distinguishes the RU approach from the NI alternative. It’s unclear how a methodology in which contextual facts constitute the entire justification for any substantive principles would constitute a normative exercise at all. Drawing substantive content entirely from contextual facts is a descriptive task.

The appropriate response to this objection is to acknowledge that ‘the repertoire of social forms relevant to justice is constrained in various ways. Societies cannot shape these forms in any way that they please if they are going to qualify as just’ (ibid., p. 22). To qualify as just, then, principles must be sensitive to some universal considered convictions in addition to their contextual grounds.

2.3.4 Limited Content Contextualism

---

\(^{33}\) Essentially, this is Gill’s (1987) argument. Others disagree and argue that Walzer’s account is contextualist in a sense that matches the LPC features. Interpreting Walzer’s account is not the focus here, and regardless of whether it strictly meets the criterion of LPC, it serves as a familiar example of how such an approach might, in theory, work.
David Miller’s approach in *Principles of Social Justice* is an illustration of a contextualist theory that avoids the criticism I pressed against LPC by ruling out the possibility that contextual facts are sufficient grounds for substantive principles. His view demonstrates one way in which a substantive principle can have both contextual and non-contextual grounds. In *Principles of Social Justice*, contexts are defined by ‘reference to the kind of social relationship that exists among the parties between whom justice is to be done’ (2002, p. 11). Membership in such a relationship represents the kind of non-universal shared circumstance that defines a contextual fact. These relationships can be more easily identified and analysed by external parties than Walzer’s shared understandings. In *Principles of Social Justice*, Miller claims that ‘empirical evidence should play a significant role in justifying a normative theory of justice’ (2001, p. 51) and that we should use ‘political theory’ (context-independent moral principles) to make sense of the empirical data. These roles for political theory and empirical data demonstrate the way in which both can feature in the grounds of a substantive principle. ‘On the one hand, the theory allows us to clarify the findings of empirical research by drawing distinctions between different principles of justice and between justice and other social values, while on the other hand the evidence helps to confirm the theory by revealing which principles people do in fact subscribe to’ (ibid.). It is only when both elements are present that theorists can use contextual facts to ground substantive principles.

LCC takes the following (general) form. The grounds of all fundamental principles in a theory of justice are partially contextual and partially non-contextual. So, a distributive principle could be justified by its coherence with both citizens’ views and universal constraints on permissible principles. I call

34 We might consider whether the nature of particular contextual facts informs the extent of their influence on substantive principles. For example, since Miller’s subject matter is observable and analysable from those external to the practices more easily, we might think that the role of these contextual facts can and should be different to Walzer’s more subjective account. I defer discussion of these ideas to section five.

35 Miller’s later work indicates that his view doesn’t require strict adherence to the idea that all principles necessarily have some contextual grounds. At the end of *Two Ways to Think about Justice* he acknowledges that there may be some context-independent principles of justice. As such, his earlier work is a more plausible example of this approach.
this view Limited Content Contextualism (LCC) because contextual facts’ influence on principles’ grounds is limited by the inclusion of the non-contextual element. This justificatory claim conflicts with my own because it makes contextual facts a necessary feature of all substantive principles’ grounds. It shares this feature with FC. On my view context forms a part of the grounds for only some substantive principles (those where the non-contextual element yields indeterminate principles).

There are two principle reasons why theorists must reject the idea that contextual grounds are necessary in the justification of every substantive principle. If this necessity rule is false, then LCC and FC are implausible approaches. First, when both contextual and non-contextual elements are present in a principle’s grounds, the non-contextual element will sometimes dominate the contextual element and make it redundant. If the contextual element can sometimes be redundant, then there’s no need to demand that it be a necessary feature of all principles’ grounds. The non-contextual element dominates when it yields determinate guidance that does not vary according to any contextual facts. Principles that respond to a universal deontological constraint, for example not killing innocent people, can never be overruled by contextual grounds, and so requiring contextual grounds in support of them is unnecessary and implausible. This possibility is sufficient to show that we should reject a requirement for all principles to have both contextual and non-contextual grounds and should instead favour weak contextualism.

The second problem with making contextual grounds a necessary condition of justification is that there may be cases in which they conflict with the universal grounds. Take a non-contextually grounded principle like equality of access to education, for example, and imagine that the social meanings of goods, or existing practices, or public opinion (or whichever contextual facts) conflicts with its recommendation and instead supports a principle of merit when it comes to education access. How should the theorist proceed? It’s seemingly impossible to adopt a distributive principle for education that draws support from both types of grounds in this case. This is a problem for RU theorists in particular because their approach requires them to develop determinate
principles of justice so that they can show them to be ultimately implementable. As such, theorists have a second sufficient reason to reject the requirement that all substantive principles of justice have contextualist grounds.

Given these difficulties with stronger variants of contextualism, theorists should favour weak contextualism. On this view contextualism merely requires that contextual facts may form part of the grounds for some substantive principles and others may have entirely non-contextual grounding. As such, they reject the universalist claim discussed in the previous section that says that contextual facts can never contribute to a fundamental principle’s grounds. Allowing universal and both contextual and universal grounding of principles of justice avoids the objections pressed at other versions of contextualism.

Weak contextualism: i) does not unnecessarily make contextual grounds a necessary condition of all principles’ justification (contrary to LCC and FC) and ii) does not risk objectionable principles by making contextual facts sufficient grounds (contrary to LPC and FC).

2.4 Existing RU Theory and Contextualism

Having defended my claim that RU theorists ought to adopt weak contextualism, I now discuss some examples of existing RU theories that accommodate justificatory principles that reflect this. Justificatory principles specify the grounds of fundamental principles of justice and I’ve argued that in RU theories at least one must be contextualist-in-content. This constrains RU theories by ruling out the possibility of only universal justificatory principles and making conceptions of justice sensitive to more than feasibility constraints alone. Although RU theorists design principles for an idealised society, weak contextualism means that the features of that utopia must include some contextual facts, rather than only universal facts that inform feasibility. In this way, weak contextualism limits the idealisations theorists are permitted to make.
Existing RU theories have different statuses with regards to weak contextualism. During the argument in section two, I pointed out that theorists like Wall and Rawls take context sensitivity to be necessary to avoid indeterminacy. I also pointed out some RU theorists who adopt a universal approach to justice and would reject my claim. Others may have under-theorised the details of their approach so that it’s unclear whether contextualism is an essential, or even optional, feature. In this section I say more about two existing methodological approaches that embrace weak contextualism, constructivism and practice dependency.

2.4.1 Constructivists

Constructivism is the view that substantive principles of justice are to be constructed from the results of appropriate procedures. Constructivists can easily incorporate weak contextualism into their approach by including contextual facts in the procedure of construction so that those facts form a part of the grounds of the resultant principles. Some existing constructivists adopt weak contextualism, although doing so is not conventionally understood to be a necessary feature of their approach. Here, I discuss two particular constructivist methodologies and show how they are weakly contextualist.

Onora O’Neill’s constructivist procedure requires *judgement* to interpret the concrete meanings of abstract moral principles. ‘Her model, despite replying on universal principles, does not require uniform action and is, on the contrary, particularly context sensitive, since the combination of extremely general principles with a fine grained exercise of judgement enables us to assess ‘the specific constructions of justice which, quite rightly, take different forms at different times and places’ (O’Neill, 1996, pp. 173-4)” (Ronzoni, 2010, p. 90). So long as we understand attention to the ‘specific constructions of justice’ as something relevantly similar to attention to the fundamental principles of justice, then it is clear how O’Neill’s approach is an example of a weakly contextualist view. In order to produce specific constructions, O’Neill relies

---

36 O’Neill herself would reject this label. For her, principles of justice are universal, but they are also ‘abstract, and therefore indeterminate’ (ibid.). This reveals a terminological dispute in which O’Neill might reject one of two claims I made in section two, that principles of justice
on contextual facts such as the specific situation, diverse cultural codes, and acceptability by those actually involved (ibid.). These are examples of contextual facts that vary across particular idealised societies.

Rawls’s (1999) constructivist procedure famously includes the use of the original position to construct fundamental principles of justice. The way that Rawls frames the use of this procedure reveals how its results are sensitive to the context they are intended to govern. In this way, his method is an example of weak contextualism. This oft overlooked contextual feature of Rawls’s approach is apparent via his distinction between the general and special conceptions of justice. ‘The General Conception applies whenever a certain basic level of material well-being, necessary to the maintenance of effective equal wealth, is unavailable. Whenever appropriate conditions are reached, however, the Special Conception, consisting of two principles, applies’ (Daniels, 1989, pp. xlvi-ii). Without the appropriate circumstances, the priority of liberty in Rawls’s special conception is not justified. The circumstances (a level of material well-being necessary to maintain effective equal wealth) are contextual facts and Rawls’s sensitivity to them makes his approach an example of weak contextualism. They form a part of the grounds that justify the principle of basic liberty and its priority in the special conception. Rawls acknowledges this explicitly when he says ‘it is difficult, and perhaps impossible, to give a complete specification of these liberties independently from the particular circumstances—social, economic, and technological—of a given society’ (1999, p. 54) and ‘There are surely circumstances in which they [the two principles of the special conception] fail [to secure justice]’ (ibid., p. 55). If his two principles fail to secure justice in certain circumstances, then it must be secured by alternative substantive principles.

This isn’t to say that Rawls’s principles are, in fact, universal but only apply in particular circumstances. As we saw in section two, such a view is compatible with the universalist approach because it permits proponents to maintain the

(for RU theorists at least) must be determinate, or that determinate formulations of moral principles should be called fundamental principles of justice. Her view is compatible with the RU approach so long as this terminological issue is overlooked.
claim that the grounds of fundamental principles are entirely universal. For this view to be plausible, the special conception of justice as fairness would have to be understood to be derivative rather than fundamental. This is implausible for two reasons. First, it doesn’t match the way Rawls regards the results of the original position. Second, and more importantly, the general (universal) conception is insufficiently determinate to constitute an RU conception of justice. As such, we can regard the principles of the special conception as fundamental. Rather than the application of the general and special conceptions drawing on contextual facts, these facts justify particular substantive principles over alternatives. In later work, Rawls was quite explicit that his aim was not to produce universal principles of justice, but rather those suitable for the specified context of a modern liberal democracy (2005, p. xxxix). He concentrates ‘specifically on the problem of a shared political conception in the wake of the history of European religious conflict’ (Räikkä, 1998, p. 271). This approach allows contexts to influence the grounds available for Rawls to defend his principles. So much is clear via his use of an overlapping consensus and a shared public political culture in the arguments in favour of justice as fairness. I consider Rawls’s later work at length in the following chapter.

2.4.2 Practice Dependency Theorists
As we have seen already in this chapter, practice dependency is a form of contextualism. Practice dependency is the view that theorists must interpret the point of the specific practice they aim to regulate when developing principles of justice. The practice dependent approach is compatible with the weak branch of contextualism that I defend.

It is clear from the previous discussion of practice dependency how proponents of this view incorporate contextual facts into their reasoning when developing substantive principles of justice. They interpret the point and purpose of practices and use these contextual facts to formulate the principles

---

37 The general conception says that all social goods should be distributed so that they benefit the worst-off [Daniels, 1989, p. xlv]. Such a principle is indeterminate in many cases where the benefits/losses to the worst off are the same amongst options. It’s also unclear that all matters of justice can be reduced to a concern with one distributive principle of social goods.
of justice that govern them. What may not be clear is how these theorists remain loyal to weak contextualism by also including non-contextual features into substantive principles’ grounds. We may consider what Sangiovanni (2008) calls the institutionalist branch of practice dependency for an illustration of how this might work. Institutionalists make the justificatory claim that the purpose of practices must be recognised in the substantive principles that govern them and that ‘the normative force of principles of justice derives directly from the first-order arguments supporting these principles’ (ibid., p. 146). A justificatory principle that reflects this would take something like the following form. ‘Conceptions of justice are grounded by independent normative standards but their justification is conditioned by the point and purpose of the practices they are intended to govern’. The appeal to independent normative grounds requires an interpretive stage of construction in which interpreters apply their own standards of value in order to ‘make sense of the point and purpose of a practice or activity, or why those engaged in it continue to affirm it (if they do)’ (ibid., p. 149). Interpreters endorse these standards of value independently of the institutions (context) they are considering. They then consider the contextual facts describing the point and purpose of the relevant practices in order to ‘establish the parameters and fixed points which a full blown conception of justice must take into account’ (ibid.). As such, the institutionalist branch of practice dependency explicitly utilises both contextual and non-contextual grounds in substantive principles’ justification and is compatible with weak contextualism.

Of course, RU theorists may adopt alternatives to constructivist or practice dependent approaches. The purpose of this section has been to illustrate how weak contextualism may be adopted given some familiar and established approaches in the literature. These existing accounts demonstrate that weak contextualism is accepted by some as a plausible approach in the literature. My contribution thus-far has been to demonstrate why all RU theorists should adopt such an approach.
2.5 Particular Contextual Facts

So far I have established the requirement that RU theorists must adopt contextual facts as the grounds of some substantive principles of justice in order to ensure determinate principles of justice and loyalty to the approach. The arguments in favour of weak contextualism do not rely on a particular conception or source of contextual facts, but rather their contextual nature itself. I’ve defined contextual facts as those that are non-universal and reflective of the shared characteristics of particular communities. The definition of contextual facts leaves open the possibility of theorists appealing to a wide variety of kinds of facts in order to meet the weak contextualism requirement. This variety includes things such as facts about the economic situation, natural resources, existing institutions, cultural beliefs, infrastructure, and history. The conclusion I have established so far in this chapter is that some substantive RU principles of justice ought to be sensitive to some contextual facts. Of course, the appeal of this claim will vary depending on the details of the contextual facts incorporated to meet it. Many of the kinds of facts listed above are, in my view, plausible interpretations of contextual facts that RU theorists could adopt to meet the weak contextualism requirement. It might be the case, however, that some are more effective means to avoiding indeterminacy than others. In this section of the chapter, I investigate the idea that people’s views about justice are a particular kind of contextual fact that ought to ground some substantive RU principles of justice.

To adhere to the requirement, theorists must include some real-world contextual facts in the description of the idealised society to which their principles apply, as well as those that inform the uncontroversial conception of feasibility. This allows these facts to feature in the grounds of principles designed to govern these societies. RU theorists also use idealisations and this makes identifying appropriate contextual facts more difficult (since there are some that can be legitimately ignored). Each claim that RU theorists should/shouldn’t include a particular kind of contextual fact into the grounds of their substantive principles of justice requires particular arguments. Some contextual facts might overlap with others in a way that complicates these
arguments (for example taking practices as the relevant kind of fact might include both historical data about, and people’s attitudes towards, those practices). These arguments need to establish why including particular contextual facts is important or helpful in virtue of their type. Then theorists can say, for example, that principles should be sensitive to contextual facts about cultural practices because of their status as cultural practices.

It seems to me that the content and success of these arguments will often depend on the reasons particular theorists adopt the RU approach over alternatives. If, for example, a theorist adopts the RU approach because she believes it to be a valuable resource to her ultimate aim of developing claims of justice that we make in the UK, today, then she might take constitutional features of the UK today to be the relevant contextual facts which ground her principles. On the other hand, if a theorist adopts the RU approach because she takes it to be the best approach for accessing the truth about justice, then we might think that only the most firmly, objectively-established, contextual facts (say geographical and historical facts) ought to be relevant.\(^38\) Regardless of the plausibility of suggestions like these, my account of the RU approach does not include a claim about why theorists ought to adopt it, and so arguments of this nature are beyond the scope of the thesis. Arguments that support sensitivity to a particular kind of contextual fact regardless of theorists’ reasons in favour of the approach are, however, relevant.

2.5.1 Weak Contextualism and People’s Views
In this section I want to consider briefly the idea that the views of those to whom principles are to apply should constitute part of the contextual facts that justify some substantive RU principles. I’m not defending the suggestion that people’s views are the only plausible candidates to meet the weak contextualism requirement, but rather that they might be essential, given certain substantive commitments. The content of a group of people’s views about justice (or morality more generally) are a non-universal feature of society that describes an

\(^{38}\) Recall that the option of designing principles for a utopia characterised by only universal facts isn’t available to RU theorists since doing so would mean that their principles would be insufficiently determinate.
element of those peoples’ shared circumstances (I discuss examples of the particular forms these could take shortly). Theorists may base their idealised society on a particular context and record the views of those within it thusly. As such, sensitivity to people’s views could be a necessary or sufficient condition of meeting the weak contextualism requirement. Miller describes the intuitive appeal of such an idea when he says that ‘I assume that one test on an acceptable normative theory is that it should be possible to make a reasonable case for the theory to those to whom it is to apply’ (1992, p. 588). This doesn’t appear to be a widely embraced intuition, regardless of whether it is shared. At this stage I’m treating the idea that sensitivity to people’s views is a necessary condition of meeting the weak contextualism requirement as a suggestion. As such, I don’t conclusively defend it, but rather suggest one reason in favour of it. I return to the idea and test its plausibility once again in each of the remaining chapters of the thesis (that identify additional methodological commitments of the RU approach). By the end of the thesis, substantive principles’ sensitivity to people’s views will be established as a unifying theme of the more stringent methodological requirements more conclusively, with support drawn from each chapter.

The way in which actual people’s views about justice could constitute a contextual fact varies. Suppose, for example, that 80% of Britons claim to believe the following: ‘how much income someone earns should not be influenced by the social class of his or her parents’. We might think that substantive principles ought to reflect, respond to, or cohere with such statements if those facts are to form a part of their grounds. Alternatively, people’s views could constitute contextual facts less coarsely via social meanings of goods or practices, pre-theoretical beliefs expressed via actions, support or compliance with institutions, the relations they exhibit with one another, or what Rawls (2005) calls the public political culture, for example.39

39 A difficulty arises when there is a distinction between what people think and what they say they think. This difficulty invites us to consider which of these accounts of people’s views is the most appropriate, given our reasons for doing so. Another problem is a lack of coherence amongst their views. To solve this problem, we need a deliberative procedure. Such procedures, and a refined idea of which facts constitute people’s views, are discussed when I reintroduce this idea in later chapters.

101
We can also understand practice dependency to be a version of sensitivity to people’s views if the point and purpose of practices is taken from the views of its participants. People’s views could feature in a practice dependent account that took the grounds of justice to be shaped by either societal cultures (in-line with the conventionalists’ view (Sangivanni, 2008, p. 138)) or institutional form (in line with the institutionalists’ view (ibid.)). Societal cultures reflect aspects of its member’s lives across the full range of human activities including both the public and private sphere (Kymlicka, 1996, p. 76). Societal cultures are defined by language and history (ibid., p. 77), both of which can plausibly be said to include a role for people’s views. With regards to the other strand of practice dependency, institutional form is partly constituted by people’s views when those institutions are formed and regulated democratically. For the remainder of the chapter I’m considering any interpretation of contextual facts that describes an account of people’s views about justice. If we conclude that RU theorists should accept this proposal and include this kind of contextual fact into some substantive principles’ grounds, then we can consider separately how, exactly, the requirement should be met. Since the proposal is only speculative at this stage, I return to this issue later in the thesis.

The reason that RU theorists ought to ground substantive principles in people’s views is because doing so is a requirement of treating those people as equals. ‘Directly enforcing a partisan conception of justice on society would be inconsistent with equal respect, because it would treat holders of competing reasonable views as inferiors’ (Valentini, 2012, p. 600). This argument is a claim about the implications of substantive commitments of RU theorists, and isn’t a methodological claim about people’s views being the best way to meet the weak contextualism requirement of the approach. RU theorists must include some conception of the views of those people to whom their principles are to apply, in order to treat them as equals.

I’m assuming rather than defending the claim that all RU theorists are committed to the substantive requirement to treat individuals equally. This commitment is pre-theoretical and universally grounded, and as such does not depend on theorists having adopted a particular approach to justice. It’s only
when we consider one of the many conceptions of equality, or particular requirements of the commitment, that we encounter more controversial claims that require defending. For the speculative purposes of this section, I adopt the familiar Rawlsian conception of equality. Many conceptions of equality ground the requirement to treat people equally in the idea that they share some capacity that warrants such treatment. This capacity is often identified to be rationality, but Rawls made a famous addition by acknowledging that *reasonableness*—our capacity for a sense of justice—is something else we possess in virtue of which we are equals (2005, p. 19).

The content of people’s capacities for a sense of justice can be ascertained through their views about justice and the only way to treat them equally with respect to this is by allowing their views to feature in the grounds of their governing substantive principles. If theorists fail to do this, then they include only their own capacity for a sense of justice into the methodology. Ignoring others’ capacity implies that the theorist considers her perspective to be epistemically superior. It reflects the sentiment ‘I know that you and I both have a capacity for a sense of justice and in virtue of that we are equal, but I don’t respect your sense of justice as much as my own and thus can see no reason to consult you when deliberating over substantive principles of justice’. To respect citizens’ status as rational agents requires their reasons to be heard (Valentini, 2013, p. 195).

In addition to this, the social and prescriptive nature of principles of justice makes other people’s capacities for a sense of justice relevant. The reason that theorists aren’t obliged to consult the judgements of all possible citizens, and what makes their views contextual, rather than universal, facts to be acknowledged, is that the substantive principles they’re considering aren’t prescriptive for all people. If theorists treat people’s views as contextual facts, then clearly not all people’s views are relevant. The social and prescriptive nature of principles of justice explains why we should regard them as contextual in this way. Principles are social because they guide the actions of individuals living together and cooperating with one another, and form a part of those individuals’ shared circumstance. Principles are only social and
prescriptive within each context to which they are to apply. Indeed for Rawls, this social nature of justice informs its content and the task of a conception of justice is to specify institutions and social arrangements that meet the requirements established by citizens’ shared understandings; a conception of justice is only grounded when it accomplishes this task. ‘Whenever a sufficient basis for agreement among citizens is not presently known, or recognized, the task of justifying a conception of justice becomes: how can people settle on a conception of justice, to serve this social role, that is (most) reasonable for them in virtue of how they conceive of their persons and construe the general features of social cooperation among persons so regarded (Rawls, 1990, p. 517). Sangiovanni acknowledges this aspect of Rawls’s approach, too. ‘In a nutshell, his [Rawls’s] view is that appealing to interpretive facts about the ‘history and traditions embedded in our public life’—including conceptions of society and person—is necessary (but not sufficient) for the justification of a conception of justice because of the social role that a conception of justice is meant to play for us here and now” (Sangiovanni, 2008, p. 151).

This point about the social nature of principles of justice providing contextual, rather than universal, reasons to include people’s views is not to say that theorists lack a duty to treat everyone equally. Whether or not such a duty exists is a separate question. If it does, then it’s unclear that it entails the same requirements as those theorists have towards those to whom the conception of justice applies.

Theorists may object that using people’s views about justice to meet the weak contextualism requirement undermines the reasoning in favour of idealisations being a part of their approach. We saw in the previous chapter that proponents of RU theory hope idealisations will help them to avoid false consciousness and status quo bias. Adopting people’s views as the facts of weak contextualism could interfere with this desire because the content of their views is likely to be blighted by these very difficulties. This should not, however, concern RU theorists to the point where they exclude people’s views. Part of the intention of including both contextual and non-contextual grounds in the weak contextualist approach is to mitigate problems found in
contextually-grounded content. In cases where people’s views are subject to false consciousness, if they believe a tax cut for the super-rich to be justifiable in a context where it clearly is not, for example, then a non-contextual commitment, that permits only moderate inequality of income, can prevent the tax cut. A principle of moderate inequality is still indeterminate, and in this sense requires some contextual grounds, but at the same time constrains the influence of contextual data. The details of such matters can be worked out in particular methodologies. I return to discuss some of them at the end of subsequent chapters.

Once again, the idea that RU theorists ought to adopt people’s views about justice as the preferred conceptual fact is only tentatively suggested here. The suggestion is spurned by the idea that in order to treat those to whom theorists are prescribing principles as equals, they must somehow include their capacity for a sense of justice into the approach. I return to this idea at the end of the next chapter.

**Conclusion**

Contextualism is the view that the grounds of some substantive principles of justice can be given by contextual facts—facts that represent the non-universal aspects of the shared circumstances of a particular community. In this chapter, I defended the claim that RU theorists ought to adopt weak contextualism.

The reasoning in support of this claim draws on reasoning introduced in the previous chapter. We saw there that RU theorists make their principles sensitive to feasibility constraints because of their core commitment to producing ‘ultimately implementable’ principles of justice. Here, I demonstrated that determinacy is a second necessary condition of this requirement. Without determinacy, theorists cannot plausibly show fundamental principles to be ultimately implementable because their requirements are unclear. In section two, I showed that allowing contextual facts to ground substantive principles of justice ensures that they are sufficiently determinate. As such, a methodological requirement of the RU
approach is to reject the universalist claim that contextual facts can never ground fundamental principles of justice. Moreover, weak contextualism is the only plausible methodological approach by which they can meet this requirement.

Until the last section of the chapter, I refrained from specifying which kind of contextual facts RU theorists ought to feature in their grounds. The argument in favour of contextualism does not depend on a particular kind of fact. In the final section, I gave a brief argument in support of people’s views about justice being the kind of contextual fact that should be included in theorists’ deliberations. I return to this idea in the chapters that follow.
3. Stability

Introduction

In this chapter I argue that RU conceptions of justice must be stable for the right reasons.\textsuperscript{40} I call this the stability requirement. A conception of justice is stable for the right reasons when it can order society with institutional support whilst treating citizens as free and equal. In this case, a conception is stable for the right reasons when ‘abiding by liberal basic laws is not a citizen’s second-best compromise in the face of the power of others, but each citizen's first-best option given their own beliefs’ (Wenar, 2013). A conception can meet this requirement if it is \textit{publicly justified}. RU conceptions must be stable because of theorists’ commitment to producing ultimately implementable principles of justice. This concern requires RU theorists to hold their conceptions of justice accountable to another standard of justification (stability), and thus to adopt a justificatory principle to reflect this. I demonstrate this, and why they should

\textsuperscript{40}Two clarificatory remarks are necessary here. First, the claim is that stability for the right reasons must be \textit{possible} for RU conceptions of justice. This is to say that a conception \textit{can} stably order a society, rather than that it \textit{will}, or \textit{does}. When I say that a conception is stable, I mean to say that stability is possible in this sense. Second, we are talking about the conception itself being stable. This might seem awkward since we may naturally think that it’s the conceptions recommended institutions, rather than the conception itself; that is or isn’t stable. However, it’s possible that various institutional alternatives could uphold a particular conception’s recommendations. What we’re interested in is whether there is something in the nature of the conception itself that prevents institutional support and citizens' compliance in this way.
accept the qualifier that principles be stable for the right reasons, in section one of the chapter.

In the sections that follow, I consider the methodological implications of the stability requirement. Once we accept the idea that stability is a necessary feature of RU conceptions of justice, we must consider what’s necessary for a theorist to show this to be the case. In section two I argue that the only way to secure stability is via public justification, and dismiss the alternative suggestion that congruence between citizens’ ideas of the right and the good is necessary. I then investigate different conceptions of public justification in order to give determinate guidance on the way in which a conception can meet the stability requirement. Broadly, these different conceptions fall into two categories: consensus and convergence. Although the details vary, both views require conceptions of justice to be supportable by citizens living under it for that conception to be publicly justifiable. So, RU theorists must develop an account of motivational constraints—what they can reasonably say their idealised citizenry will and will not be motivated by—in order to meet the stability requirement. RU theorists ought to reject the convergence view and adopt consensus. I also offer a revised account of consensus. Theorists should prefer this view because it permits broader and more meaningful public justification. Public justification according to the conventional interpretation of consensus is, however, sufficient.

Finally, in section four, I consider the compatibility of the stability requirement with the conclusions I draw in the previous chapter. At the end of chapter two I suggested that people’s views about justice might be an appropriate kind of contextual fact to meet the contextualism requirement. In this chapter, I show that suggestion to be compatible with the stability requirement and, as such, that RU theorists can be weak contextualists and meet the stability requirement via the same justificatory principle.
3.1 RU Conceptions of Justice Must be Stable for the Right Reasons

In this section I show first why any RU conception of justice must be stable, and then why the qualifier ‘for the right reasons’ is necessary. The stability of a conception of justice, as we have seen, refers to its capability to order society over time and generate its own support from citizens. A conception of justice is unstable when the content of its recommended principles eventually undermines institutions’ ability to order society according to them. For example, if democratic rights extended to the power to remove the equal rights the democratic procedure secures (Valentini, 2013, pp. 181-2), then a conception of justice ordered by such rights could undermine itself. The idea that RU conceptions of justice must avoid this possibility is what I call the stability requirement. The concept of stability was first given serious attention by Rawls. For him, ‘in order to be stable a conception of justice must generate its own support and the institutions to which it leads must be self-enforcing, at least under reasonably favourable conditions’ (2001, pp. 124-5). By constraining the requirement to ‘reasonably favourable conditions’ we can see that Rawls intended the requirement to apply within an idealised society—the kind with which RU theorists are concerned. As we saw in chapter one, RU theorists use idealisations to map the society for which they are developing principles, and the limits of these idealisations are given by feasibility constraints and sensitivity to some contextual facts (as I argued in chapter two). Using idealisations and acknowledging that they should be constrained in this way distinguishes the RU account from alternatives.

Before investigating how to determine whether a conception of justice meets the stability requirement, I first outline why RU theorists should accept it. Having done this, I demonstrate why we should prefer a conception of stability for the right reasons.

3.1.1 RU Conceptions of Justice Must be Stable

We may distinguish RU theories from their Utopian alternative by noting the requirement that their principles be ultimately implementable. A necessary
condition of their being ultimately implementable is their feasibility. A principle is feasible, recall, when its recommended actions can be completed in those possible worlds subject to the circumstances of justice (with a reasonably likely chance of success). If actions are unfeasible, then the principle cannot be ultimately implemented and thus fails to be action-guiding in the RU sense. In the second chapter, I argued that context-sensitivity also emerges from the concern to develop these kinds of action guiding principles.

Here, I argue that RU theorists should have a similar concern for stability. That is, stability should be recognised as an additional requirement of producing ultimately implementable conceptions of justice. Rawls most famously endorses the idea that stability is a necessary feature. According to him, ‘if a theory is unstable then it is not a satisfactory conception of justice and must in some way be revised’ (2005, p. 141). Stability is a vital aspect of a conception of justice simply because it is unsatisfactory for the institutions that support a conception of justice to be unmaintainable. If people’s ordinary compliance with the principles results in their downfall or undermining, then we should prefer an alternative.

Classical Utilitarianism is a good example of a conception of justice that could be unstable in this way. A conception that requires individuals to act in such a way to maximise aggregate utility might be unstable for a number of reasons, the two most significant being found in well-rehearsed objections to the approach. First, complicated calculations to determine which action yields the highest utility score would undermine participants’ ability to meet the requirements of the approach. This gives us a stability-based reason to reject the conception of utilitarianism in favour of rule or indirect utilitarianism. Second, the kind of impartiality it requires is beyond the motivational capacities of most humans making them unable to comply with the conception’s demands and undermine its ability to guide their actions.

RU theorists should require the stability of their conception of justice for the same reason they require its feasibility. A conception of justice is not ‘implementable’ in any meaningful sense if it is not stable. Imagine a conception of justice that was extremely demanding on natural resources; it
requires they are generously distributed amongst citizens. This might be ultimately implementable in the sense that we could arrange such a distribution at some point, but there wouldn’t be enough left to sustain it for more than one generation. Intuitively, we don’t want to say that such a distribution is implementable if we mean that it is implementable for a (comparatively) short time only (and can only ever be so). The question of stability concerns whether or not a conception of justice can order a society over time. If a conception is unstable, then it can’t order society in this way and RU theorists should not regard its principles to be ultimately implementable. As such, an unstable conception of justice can’t provide action-guiding recommendations in the RU sense. Recall that RU principles are designed to order social arrangements. If the very nature of a conception of justice prevents it from ordering society over time, then its principles are not guiding-action for the constituents of that society.

Moreover, ignoring the stability requirement is implausible given the methodological commitments I have established so far, in particular the commitment to feasibility constraints. Theorists should not accept the reasoning in favour of feasibility constraints but reject those reasons when it comes to stability. If one is important, then the other must be too. RU theorists value feasibility because they view the aim of justice to be to provide action-guiding recommendations for a moderately idealised society. It’s implausible for RU theorists to be concerned by infeasibility undermining the action-guiding nature of their principles, but not to be similarly concerned by instability doing the same thing.

Although the reasoning in favour of feasibility constraints informs the reasoning in favour of the stability requirement, fulfilling the former does not guarantee that the requirements of the latter are met. The reasoning that motivates a concern with stability does not imply that unstable principles will also be infeasible. Although some theorists disagree that the two concepts should be separated in this way, doing so is necessary to preserve the uncontroversial conception of feasibility I introduced in the first chapter that we may attribute to all existing RU theorists. Feasibility is an aspect of the RU
methodology that existing proponents accept, but stability is much more controversial or at least appears to be by the relatively little mention of it in the literature. Although I left open the possibility for RU theorists to supplement the feasibility account, doing so here would be to misunderstand the nature of the stability claim. On my view, the two attributes are conceptually distinct, despite the convention expressed by the following two theorists.

Some might prefer to leave stability out of a conceptual account of feasibility, and in that case it could be repackaged as part of the desirability of some action. It is undesirable to put our efforts into pursuing an outcome that won’t last (unless it is a transitional outcome). But it seems to me that the recommendations of political theories come with an implicit or explicit temporal scope, so that when we ask whether an outcome is feasible, what we’re really asking is whether an outcome, for a given length of time, is feasible. The answer will be ‘yes’ if the outcome, for that given length of time, meets the conditions for binary feasibility (Lawford-Smith, 2013, p. FN. 20)

Similarly, Raikka.

The point is that the social ideals are infeasible only if there are strong constraints that limit the implementation of the ideals. Strong constraints make it impossible to implement the ideals or, once the ideals are in place, to maintain them (the stability problem). FN: so, the question of stability is a sub-question of the feasibility question. Feasibility bears on the durability or stability of a regime as well as its accessibility and possibility. One possible reason why an ideal is not feasible is that it is not stable (1998, p. 32)

Essentially, feasibility is concerned with whether it is possible, given the world roughly as we know it, for a conception of justice to be implemented. The qualities that ensure this make a conception sensitive to facts about the laws of
nature and the circumstances of justice, as we have seen. There’s no reason to presume that such sensitivity will ensure that a conception can order society over time. There are a number of additional factors—most notably citizens’ motivations—that contribute to stability but not feasibility. What citizens are capable of being motivated by does not affect feasibility because it only requires theorists to consider implementing a conception at a particular point in time, and thus merely whether it is possible for citizens to comply with requirements (including whether it’s possible for them to be motivated as such). Stability, on the other hand, is about whether theorists can reasonably claim that citizens will be motivated to comply over time. Clearly this requirement imposes different and likely more substantial constraints. It’s feasible for me to willingly babysit Charlie, a difficult 2 year old, but it’s a separate claim altogether to reasonably claim that I would be motivated to continually do it over time.

Another way to recognise the distinction is by considering the view of a Utopian theorist who is not concerned with feasibility. Such a theorist might only be concerned with designing principles to govern a far-off abstract society that isn’t necessarily a possible version of the world that we’re in. It’s coherent for that theorist to think that a valuable attribute for a conception of justice is that it is stable, and that its contents don’t necessarily undermine themselves over time. Since she can endorse and establish this quality independently of feasibility, it’s clear that the two concepts are distinct.

Of course, there are other qualities by which we may assess competing conceptions of justice. This requirement states that RU conceptions must be stable; if they are not, then they cannot be RU conceptions of justice. Once the stability requirement is satisfied, theorists may assess the merits of competing theories on other grounds, according to their particular approaches.

3.1.2 RU Conceptions of Justice Must be Stable for the Right Reasons

Once we accept the idea that RU conceptions of justice should be stable, we must consider what’s needed for a theorist to show this requirement is met. I defend the idea that theorists must show their conception of justice to be stable for the right reasons. In short, a conception of justice that is stable for the
right reasons ‘must be able to win support by addressing the reasoning of each citizen rather than by bringing in sanctions to ensure citizens act in accordance with it’ (Rawls, 2005, p. 143). So, to achieve stability for the right reasons we want citizens’ actions to be moved by their reasoning about justice, and for them to conclude that supporting just institutions is the appropriate action for them to take. There are two problems with stability that is not achieved in this way.

Stability achieved by the use of brute force, and brute force alone, is not achieved ‘for the right reasons’. If a conception of justice only continues to order society’s institutions because of the state’s powerful and coercive enforcement of it, then stability depends on brute force. It is this kind of stability that concerned classical theorists like Plato and Hobbes (Weithman, 2011, pp. 44-45) (Freeman, 2003, pp. 278-9). We ought to reject the notion that these circumstances ensure any kind of valuable stability because they necessarily undermine citizens’ status as free and equal. 41 This substantive implication is unacceptable and so stability achieved in this way does not fulfil the requirement in the spirit in which it is intended. Whether or not a conception can order a society in this way does not depend on any features of that conception. As such, it does not make sense to call this kind of stability a valuable attribute of a conception. Similarly, if a state were to force everyone into being altruistic, we wouldn’t say a conception of justice that required unlimited altruism became stable as a result.

Maintaining the social order of a conception of justice by force or the threat of force undermines citizens’ status as free by preventing them from making their own decisions about how to live, and by making them unfree to pursue certain courses of action. It undermines citizens’ status as equal by ignoring their equal right to self-govern, and treating them as less equal than the law-makers and enforcers. This is not to say that all coercion is impermissible; there are some cases where coercing citizens is compatible with the values of freedom and equality.

41 I’m assuming that commitment to some conception of each of these values is a substantive feature of any plausible RU theory.
equality. For these reasons, this kind of stability cannot be a valuable attribute of a conception of justice. Neither can it be said to define an RU conception of justice because it is incompatible with substantive demands of justice that I take any plausible version of the RU approach to be committed to. RU theorists should be interested in finding a conception that can order society over time without the need for such impermissible coercion. Then, we can understand stability as a quality of a conception of justice that counts in favour of it justly governing the constituents of a moderately idealised utopia.

One way that we can achieve stability without the impermissible use of force is via a *modus vivendi*. A modus vivendi involves citizens’ abiding by the requirements of a conception of justice because it’s in their interests given the circumstances, or because there is a balance in the forces of those competing for power (Rawls, 2005, p. 147). In this case citizens support and comply with a conception of justice because doing so is in their interest at the time. Achieving stability in this way also conflicts with stability for the right reasons. This kind of stability isn’t valuable because there is no guarantee that it will last. If the circumstances change, and the balance of power shifts so that it is no longer in citizens’ interest to support the status quo, then the institutions of the society may become unstable and the conception of justice is undermined (ibid.). The fact that there is no guarantee that it will last over time means that it does not address the very reason that RU theorists need to show their conception of justice to be stable in the first place. Even if we could design institutions that ensured the circumstances amenable to a modus vivendi endured over time (which is an unlikely possibility), this kind of stability would not be good enough.

Recall that we are demanding that conceptions of justice must be stable and that it being so somehow counts in its favour as a *conception of justice*. If the requirements of a conception of justice match the way citizens’ self-interest directs them to behave, then this is good practically speaking since it avoids the

---

42Coercion is, in fact, necessary to assure citizens that others will abide the principles of justice. Accounts of legitimacy explain in what circumstances and to what extent such coercion is permissible. This is the subject of the following chapter.
necessity for coercive force, but it’s unclear how this would count in favour of the justness of a conception. What stability should require is a conception to show that citizens comply when they are guided by justice rather than the threat of force or the immediate consequences to their interests. Rawls includes citizens’ sense of justice into the conception of stability and stipulates that it must be ‘strong enough to resist the normal tendencies to injustice. Citizens act willingly so as to give one another justice over time. Stability is secured by sufficient motivation of the appropriate kind acquired under just institutions’ (ibid., pp. 142-3).

Having citizens support a conception from within their own sense of justice is superior to the stability of a modus vivendi because it i) doesn’t rely on the threat of force, ii) doesn’t rely on happenstance, and iii) gives citizens principled reasons from within their own sense of justice to act. The last point demonstrates how stability for the right reasons matches with the RU view that the purpose of principles is to guide-action. The very purpose of principles of justice is to give citizens reasons to act that do not originate from their own self-interest. Reasons from self-interest are not reasons of justice. Because RU theorists are reasoning about principles that guide a moderately idealised utopia, they can consider this question from the perspective of citizens with a sense of justice and ask what is required of a conception in order to say that justice-motivated citizens will act in accordance with it. I’ll say more about the way citizens may be idealised by RU theorists in section 3.4. The way in which we should assess the stability of a conception is by asking whether individuals can be motivated, from within their sense of justice, to support it over time. Such support demonstrates that a conception of justice can be stable for the right reasons. This sense of stability is clearly a virtue of a conception of justice.

3.1.3 Motivation and Stability for the Right Reasons
Once we understand stability in this way, then we can see that showing a conception of justice to be stable might require more careful consideration than it first seemed. The requirement that citizens act according to their sense of justice is especially demanding when we consider the fact of pluralism; the
idea that citizens will inevitably disagree about the aims and goals that constitute part of a worthwhile life (their conception of the good) and the role that a conception of justice should play within those aims and goals. ‘As he [Rawls] came more deeply to appreciate the fact of pluralism, he came to realize that what really needs to be removed are any temptations to act against the sense of justice that arise from within any other point of view than “the point of view of justice”’ (Weithman, 2011, pp. 263-4). In order for a conception of justice to be stable in this way, citizens must willingly put aside any demands originating from their good that conflict with the requirements of justice. A big concern for Rawls, then, was whether such claims would inevitably destabilise conceptions of justice by compelling different citizens to fulfil different conflicting requirements. The potential threat of divergent ideals of the good life incompatible with a conception of justice means that theorists must make their conception sensitive to some motivational facts so that citizens can willingly act according to it. So, as Freeman says, ‘a conception of justice (like any moral conception) should be compatible with our moral and psychological capacities. It should be responsive to basic human needs and interests, at least in so far as conscientious moral agents who affirm the conception should be capable of developing appropriate attitudes enabling them to normally and regularly act upon its demands given the constraints of human nature’ (Freeman, 2009, p. 46). The question of stability for the right reasons is essentially a question of the RU theorist reasonably claiming that citizens can be motivated by their sense of justice to comply with her conception of justice, given their divergent and conflicting ideals of the good life.

Motivational constraints on conceptions of justice are a complicated matter. Partiality towards oneself and family members, for example, is widely accepted as a fundamental characteristic of living a good life and, as such, imposes a constraint on justice. There is no possible world in which humans can shed all ties with family members. Because of this, theorists may reject principles of justice that require total impartiality between family members and strangers on feasibility grounds (since this feature of humans is true in all possible worlds). Other motivational constraints, for example someone not being motivated to
fulfil a requirement of justice because she is lazy, are deemed irrelevant. Intuitively, we don’t think that Lazy Lois is exempt from a requirement to save an innocent drowning person purely on the grounds that she can’t be bothered (i.e. she lacks the motivational capacity). Our intuitions about this case remain even if we stipulate laziness to be a part of Lois’s conception of the good. Part of the difficulty here is epistemic; there’s no way of us knowing whether Lois is subject to the same motivational barriers that we assume everyone faces with regards to the demand to be impartial to strangers. There’s also no way of knowing whether laziness is a part of Lois’s fundamental commitments, properly understood. Some theorists, for example David Estlund (2011), have expressly denied the idea that motivations impose constraints on requirements of justice in this way. He thinks that in cases where they appear to, it is never their feature as a fact of human nature that gives them this characteristic. Estlund, however, is pursuing the Theory of Ideals and so his claims don’t necessarily apply to the RU approach. Motivational constraints are something that all existing RU theorists accept when they are true of all humans in any and all possible worlds. The stability requirement means that theorists must consider motivation more broadly, in terms of individuals’ conceptions of the good conflicting with the actions justice requires of them.

At this point we can reconsider the distinction between facts that inform feasibility and facts that inform stability. To determine whether a conception is feasible, I said that RU theorists ought to make their principles of justice sensitive to facts that shape the circumstances of justice and the laws of physics that shape all possible worlds. There are some motivational facts that are true of humans in all possible worlds, partiality to oneself perhaps being the most uncontroversial example, and at least one (limited altruism) shapes the circumstances of justice. Sensitivity to these facts is, however, insufficient to say that a conception is stable as well as feasible. As I have said, the possibility of citizens complying with a conception’s requirements is insufficient to show

---

43 So much is clear from the following quote: ‘what institutions a society should implement given such practical constraints can hardly be thought to deliver the content of the idea of social justice’ (ibid., p. 227). If Estlund does not consider principles’ purpose to be regulating social arrangements, then he is not committed to an approach on the ideal/non-ideal continuum.
that they could be motivated to adhere to it over time, for the right reasons. Consider a principle that required us to think carefully about the division of moral labour—how requirements of justice compete with other moral requirements we see ourselves to have—every day for 3 hours. According to our conception of feasibility, this is a plausible requirement; there are no psychological facts of human nature with which it conflicts, and it does not undermine limited altruism. It’s not plausible, however, to say that the principle can be stably maintained over time. When a conception of justice is stable it engenders in people ‘the requisite desire to act upon [a conception of justice]’ (Rawls, 1999, p. 398). The facts that inform this possibility are clearly different from those that inform a feasibility assessment. So, RU theorists must supplement their theory with justificatory and methodological principles that make conceptions sensitive to motivational facts in addition to those required by feasibility and contextualism.

Since not all humans have the same motivational obstacles in terms of the way that their conception of the good weighs against claims of justice, and we’re lacking the empirical capacity to make individual assessments for everyone, RU theorists must idealise citizens before they establish the necessary motivational barriers to requirements. This idealisation will consist in assuming that they are capable of being motivated by their sense of justice, and may include additional features too. As long as the idealisation does not undermine what humans are reasonably capable of, then principles developed on the basis of it can guide-action. Indeed the very aim of the RU approach is to consider a conception of justice that governs people who are as just as we (humans) can be. Once these idealisations and methodological claims are established, conceptions of justice can be tested to see whether they are stable for the right reasons. I discuss these methodological claims in the following section. If theorists can reasonably claim that citizens will be motivated to comply with a conception, despite potential conflicts with their own ends and others’, then that conception is stable for the right reasons.

44 For the source of the phrase ‘the division of moral labour’ see Scheffler (2005).
3.1.4 Existing RU Theorists and Stability

In this section, I briefly mention some existing RU theorist’s attitudes to stability. As we have seen, Rawls is the leading proponent of the view that conceptions of justice must be stable. He, like other liberals (Nagel, 1991, p. 27), takes stability for the right reasons in particular to be essential for any conception of justice that aims at practical guidance. Others, like Buchanan, acknowledge that stability is valuable, but not exclusively for the right reasons since enforcement of citizens’ individual rights can be ‘externally imposed or internally monitored’ (2000, p. 27). As such, the motivational capacities of citizens to comply are not a requirement of stability on his account. He does, however, acknowledge that conceptions of justice may be constrained by ‘possible sets of rules for personal behaviour’ (ibid.), and so we can see that some account of permissible limits of human motivation is required in his view.

Gauthier appeals to a similar concept of the sense of justice and endorses stability for the right reasons when he says that ‘an affective capacity for morality is needed if the constraints required by essential justice are to be willingly honoured. Without an emotional commitment by its members to maintain the frame-work of mutual benefit, a society could ensure its stability only by imposing on them, through processes of socialization, loyalty to a more substantive goal, which would define roles that individuals would not be free to accept or reject’ (Gauthier, 1986, p. 348).

This view contrasts with Rawls’s since Gauthier rejects the idea that it is acceptable to inculcate loyalty to a more substantive goal through socialisation. Despite the differences in reasons to impose a stability requirement, the broad acceptance of the concept should further demonstrate its appealing nature. In the following sections, we will see that the particular account of stability that I endorse for RU theorists shares many features with Rawls’s solution. There are, however, some aspects of his approach that require refinement or further discussion. Departures from Rawls’s account are identified as such.

I assume that those RU theorists who do not discuss stability do not take it to be a necessary condition of an RU conception of justice. In this sense there are
some theorists to persuade. I take the discussion in this chapter to demonstrate why RU theorists’ desire to develop ultimately implementable principles of justice commits them to accepting the stability requirement. In addition, theorists disagree over the methodological implications of this requirement. It is not my intention to develop a precise and fully-worked-out conception of stability or to specify a particular account of human motivation/psychology. Rather, I consider what is required of a conception of justice in order for theorists to show that it can comply with the stability requirement in the section that follows. From here, whenever I mention stability I will have ‘stability for the right reasons’ in mind.

3.2 Stability for the Right Reasons Requires Public Justification

In this section I discuss the implications of the stability requirement in terms of RU methodology. RU theorists must adopt a justificatory principle that specifies how they can show a conception of justice to be stable. When a conception of justice meets the requirements of such a justificatory principle, we can be confident that it is stable in the relevant sense. The requirements, as we have seen, will specify the kind of motivational facts that RU principles must be sensitive to, and those features of citizens that theorists may idealise. These justificatory principles commit RU theorists to a more determinate methodology than proponents of the approach typically acknowledge.

It’s not the case that a commitment to stability requires each and every motivational fact of citizens to be a constraint on the requirements of justice—Lazy Lois was an example of the type of motivational constraint that’s intuitively implausible. Not only are some motivational demands implausible constraints, but RU theorists needn’t be bound by all such facts because their task is not to find a set of principles that are actually stable in the real world. Rather, RU theorists just need to be able to reasonably claim that their conceptions are stable in the type of moderately idealised society for which
they are designing principles. Whether or not a conception of justice turns out to be stable in the real world is subject to a number of additional empirical considerations that should not influence RU conceptions of justice. I consider two suggestions of justificatory requirements that, when met, show that theorists can make the reasonable judgement that a conception of justice is stable. This is the appropriate standard of stability for RU theorists given the nature of their substantive principles and their core commitments. Other approaches to justice, with different core commitments, might value stability and satisfy its requirements differently. These differences are outside the scope of the thesis.

In section 3.2.1, I consider the suggestion that citizens must view conceptions to be congruent with their good in order to be stable. I reject this account because it requires RU theorists to make impermissible idealisations. In section 3.2.2 I consider the most plausible alternative: the idea that public justification is sufficient to demonstrate stability. This account shows a much more plausible compatibility between citizens’ motivational requirements and a conception of justice, and does not require any impermissible idealisations.

3.2.1 Congruence

Congruence: citizens have reasons from within their conception of the good to support, uphold, and prioritise the principles of justice.

One view states that a conception of justice must be congruent with citizens’ ideals of the good life in order to be stable. Recall that a conception of the good represents the aims and goals that each citizen sees as a part of leading a worthwhile life. To achieve congruence, it needs to be true that having, maintaining, and acting upon a sense of justice is a part of each individual’s good and that the requirements of the conception of justice are compatible with the requirements of a good and worthwhile life. The sense of justice is the individual’s capacity to understand, apply, and act from principles of justice (Rawls, 2005, p. 75). So, a society shaped by institutions that ensure citizens possess a sense of justice as a part of their good ‘not only generate its own supportive moral attitudes, but these attitudes are desirable from the standpoint of rational persons who have them when they assess their situation
independently from the constraints of justice’ (Rawls, 1999, p. 350). Here, Rawls appeals to an ideal of the good that includes pursuit of justice as part of living a valuable life. The argument was that, given conjectured psychological principles and the benefits secured to each citizen under the two principles, citizens raised in a society well ordered by Rawls's two principles would naturally develop a sense of justice and find a just life largely congruent with their various conceptions of a good life’ (Hill, 2000, p. 243). When ideals of the good require compliance with just institutions in this way, it makes following your sense of justice part of your rational plan of life and the two pursuits (the right and the good) are congruent with one another.

The view that congruence between citizens’ sense of justice and ideal of the good is necessary to ensure a stable conception of justice rests on an impermissible idealisation, and as such RU theorists must reject it. The idealisation requires theorists to assume away the fact of pluralism and imagine a society in which each citizen shares a conception of the good. The ideal of congruence depends on this idealisation because it is the only way that it can explain the possibility of a conception’s being stable for the right reasons. If citizens have divergent conceptions of the good, then they cannot possibly all be congruent with a particular conception of justice. As we have seen, the fact of pluralism means that citizens will inevitably come to disagree in this respect.

This is the case, in particular, for liberal accounts of justice since they secure the basic freedoms of thought and expression necessary for such ideas to thrive (Rawls, 2005, p. 36). The nature of humans and the burdens of judgement leads to this kind of disagreement and, absent impermissible coercion (that fails to treat citizens as free and equal), divergent ideas of the good life will always remain. Rawls was particularly concerned by reasonable disagreement. Reasonable disagreement persists over ideals of justice and the good life and is of particular interest because it is unclear that we have conclusive ways to establish whether any reasonable views are more plausible than another. For instance, ‘citizens disagree over whether abortion is morally justified, yet often neither party to this disagreement can be shown to be clearly mistaken…Of course, an anti-abortionist might find the views of a pro-choice activist obviously mistaken, but the question is whether she has grounds for so
thinking. The issue is hotly contested, and it seems dogmatic, from a liberal perspective, to deny that there is a range of reasonable positions on the matter (Valentini, 2013, p. 185). Part of the reason that this is so difficult is that the source of the disagreement, what Rawls calls the burdens of judgement, is ‘the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgement in the ordinary course of political life’ (2005, p. 56). Secondly, the hazards that impede our judgements in this way are unavoidable, and so no-one can be blamed for the disagreement resulting from insufficient consideration or something like that. As Freeman puts it, ‘the problem with congruence was that, even if it is true that justice is a highest-order good since it realizes the capacities for human agency and moral autonomy, due to the burdens of judgment many reasonable people in a well-ordered society refuse to recognize justice in this manner’ (2007, p. 33). It simply won’t be the case that citizens in any version of society resembling reality will share a particular conception of justice and a comprehensive view of the good life that prioritises it. Even if we imagined a society in which every citizen endorsed the same religion, and the good of that religion included the priority of justice, the burdens of judgement would mean that such a belief couldn’t possibly be maintained over time.

In light of reasonable pluralism, RU theorists would have to deploy infeasible idealisations to demonstrate that a conception of justice could be congruent in this way. A citizen achieves congruence when a sense of justice features prominently as a part of her ideal of a good life. To achieve this kind of congruence, RU theorists must show that each citizen’s conception of the good is shared with others’ to the extent that they agree over the priority and good of justice. A society could only be stable in this way if all citizens were substantively committed to a particular conception of justice and its priority. Given the fact of reasonable pluralism, a sufficiently conclusive defence in support of any particular conception is unavailable to RU theorists. Moreover, it’s infeasible to implement given people as they are. For it to be implementable theorists must imagine their conception to govern the constituents of an idealised society in which each citizen independently adopts the same conception of the good. This is the only way for theorists to achieve
stability via congruence and maintain their commitment to the free and equal treatment of citizens. Having done this, the theorist must then demonstrate that her particular conception of justice would be supportable from the ideal of the good that the citizens endorse. If people holding this ideal of the good would be motivated to comply with the conception of justice, and thus the institutions supporting it, then it would be stable for the right reasons.

This idealisation—to a group of citizens who all share a conception of justice and include it in their interpretation of a valuable life—is impermissible. Idealisations are only permitted as long as using them doesn’t undermine the RU action-guiding nature of principles produced on the basis of them. Principles developed on the basis of this assumption will fall prey to the guidance critique (Valentini, 2009, p. 333). They will be unable to guide-action in any plausible version of the real world where individual’s notions of the right and the good are not congruent in this way. Assuming that there is one defendable ideal of the good (the imposition of which wouldn’t violate citizens freedom or equality), or assuming that all of the members of a society can independently converge on a particular ideal of this kind is beyond what humans are reasonably capable of. It is to make the assumption that all people have the positive motivation to act according to the demands of justice because doing so is part of what makes a good and valuable life. Since the aim of RU principles is to guide action in a moderately idealised, but possible, world, this is not acceptable. As such, congruence is an unacceptable account of stability—theorists cannot meet the stability requirement by adopting it—because it violates one of the necessary aspects of the RU approach that I have already established. It is unfeasible, and the idealisations necessary to make its achievement feasible are impermissible. So, it’s not the case that infeasibility damages its stability-ensuring qualities (it would still produce a society that was stable for the right reasons), but that RU theorists must reject it on this basis.

In this section I have described congruence as the requirement that citizens share a pursuit of the good that includes their supporting and conforming to just institutions in order for the conception that guides those institutions to be stable. Understood in this way, congruence can ensure stability, but it can’t be a
necessary feature of an RU conception of justice because it cannot achieve
stability for the right reasons and be compatible with the fact of reasonable
pluralism. So, in order to demonstrate that a conception of justice can be
stable for the right reasons, RU theorists have to show that it can
simultaneously motivate citizens with divergent views of the good life.45

3.2.2 Public Justification

Rawls famously labels the idea that multiple divergent accounts of the good life
can provide citizens with reasons to comply with the demands of justice the
Overlapping Consensus. In such a consensus ideals of the good life endorse a
political conception of justice, each from its own point of view (2005, p. 134).
It resolves the problem raised by reasonable pluralism by describing how ‘all
reasonable people still can regard justice as essential to their good and thus
rational, but not because it realizes their capacities for agency and their moral
autonomy [a particular ideal of the good]’ (Freeman, 2007, p. 33). Rather than
relying on the priority of justice via a particular ideal of the good life, the
overlapping consensus relies on political values that can work as the basis of
public justification (Rawls, 2005, p. 140).

According to public justification, a conception of justice is stable if it can be
the subject of mutual agreement amongst citizens. In this sense, a theorist must
show her conception to be the subject of public agreement and this means she
must consider the justness of its requirements in the context of citizens’
reasoning addressed to one another, rather than her reasoning addressed to
citizens (or other theorists). There are numerous accounts of public
justification, with varying justificatory principles to reflect this demand, but ‘the
fundamental moral idea behind such a position is that we should not impose
arrangements, institutions, or requirements on other people on grounds that
they could reasonably reject (where reasonableness is not simply a function of
the independent rightness or wrongness of the arrangements in question, but
genuinely depends on the point of view of the individual to some extent)’

45 This ideal can be understood as a type of congruence, although it has a different nature since
agreement can come from within different sources of the good. Despite this possibility,
whenever I talk about congruence I mean the simpler idea that I have just rejected, where each
person must support share the ideal of justice as a priority good in itself.
(Nagel, 1987, p. 221). I specify the conception of public justification I prefer in the following section. Here I merely want to describe the way in which appealing to political values as the basis of public justification can secure stability in a way that congruence cannot. Public justification’s fundamental nature, as public, allows it to avoid the difficulties faced by congruence.

Like congruence, this account of stability requires citizens to have reasons to support and comply with a conception’s recommended institutions; ‘the stability of the WOS [well-ordered society—Rawls’s idealised society] depends upon all members of the WOS wanting to live up to certain ideals or conceptions of themselves that are realized only when they maintain their desire to act from the demands of the right’ (Weithman, 2011, p. 271). When a conception of justice is publicly justified, then the reasons in favour of it (demands of the right) are acceptable to all citizens without each of them sharing a particular conception of the good. So long as each citizen has a reason to support the conception, then it can order society without violating their statuses as free and equal citizens. When theorists adopt public justification they treat people as free because they are not required to adopt a particular conception of justice as a part of their ideal of a good life, leaving them free to identify it for themselves. Theorists also treat people as equals when they adopt public justification standards because particular ideals of the good life are not undermined at the expense of others.

RU theorists should prefer this approach to securing stability for the right reasons because it does not require them to make impermissible idealisations with regards to citizens’ motivations (by assuming they all share the ideal of justice as a part of the good life). Some idealisation of citizens is still necessary. As we have seen, theorists must assume that citizens possess the moral power of a sense of justice. The task given by the stability requirement is to demonstrate whether a particular conception is supportable by this sense, given pluralism about the good life. Constraints on permissible idealisations for RU theorists mean that the alternative to public justification, congruence, is unacceptable. The details of the public justification approach—what kind of justificatory principles RUs must adopt and what this means in terms of their
treatment of motivational constraints—is the subject of the following section. I argue that the consensus view is acceptable to RU theorists, but they should prefer an alternative that I call the revised consensus view.

3.3 Versions of Public Justification

In this section I take up the question of what it means for a conception of justice, or a particular proposal within a conception, to be publicly justified. If a conception is publicly justifiable, then it is stable for the right reasons. Disagreements between public justification theorists can be characterised as a dispute over whether they prefer the consensus or convergence standard of justification. I argue that RU theorists should prefer what I call the revised consensus view (or revised view, for short), but that the conventional consensus view is also acceptable. On the convergence view (which RU theorists should reject), a proposal is publicly justified when each citizen has a sufficient reason to support it. For consensus theorists, proposals are only publicly justified when each citizen has sufficient reason to support the proposal and when those sufficient reasons draw on public or shared values or ideals only (Quong, 2011, p. 264). These are public reasons (Rawls, 1997). Theorists take these reasons to be shared with other citizens when they draw on shared values in this way. Non-public reasons, on the other hand, rely on one’s acceptance of a comprehensive view of morality and the values it supports, and thus are not shared by every citizen. The revised view says that sufficient shared reasons are necessary for public justification but that non-public reasons can also contribute to their justification. In this section my aim is to show that both the consensus and revised views are acceptable to RU theorists, but that they should prefer the latter. The revised view is preferable because it permits a deeper type of justification that should be attractive to RU theorists.

The structure of the section is as follows. I start, in 3.3.1, by introducing the details of a public justification approach via two particular requirements that I take every proponent to endorse, regardless of her particular methodology. These are the sincerity requirement and the sufficiency requirement. The
sincerity requirement stipulates that citizens must be able to sincerely offer other citizens a reason in favour of a publicly justified proposal. The sufficiency requirement requires each citizen to have a sufficient reason to support the publicly justified proposal herself. In section 3.3.2, I show that convergence public justification cannot adequately meet the sincerity principle. In the two sections that follow, I introduce two concepts that proponents of the consensus view deploy in order to avoid this criticism. The first is the public political culture, which they use as a source of values from which to reason. The second is the idealisation of the constituency to reasonable citizens. I clarify how this idealisation is appropriate for RU theorists and compatible with the other requirements of their methodology. Then, in 3.3.5, I focus on the consensus view and the way in which it successfully meets both requirements of public justification. This success makes it suitable for RU theorists. Despite this suitability, I propose a revision to the consensus view in section 3.3.6. The revised view benefits from meeting the sincerity principle in the same way that conventional consensus recommends. They differ with regard to the role they permit for comprehensive reasons. The revised view acknowledges that such reasons may contribute to the public justification of a proposal, whereas proponents of the conventional view reject this. I give some reasons in favour of the revised view although I do not argue that they are sufficient to settle the case in terms of its superiority to conventional consensus.

3.3.1 Two Requirements of Public Justification

The general principle of public justification is widely endorsed. “The Public Justification Principle (PJP) states that: A coercive law L is justified in a public P if and only if each member i of P has sufficient reason(s) Ri to endorse L’ (Vallier & D’Agostino, 2014).” The following two requirements elaborate on this general principle, capturing intuitions shared by public justification liberals and recognisable in any account. First, public justification requires that each citizen has sufficient reason to support the proposal (the sufficiency requirement). The significance of this requirement should be clear. Public

---

*The coercive nature of such laws does not concern me here. It is the focus of the next chapter. Here I talk of the justifiability of proposals, or conceptions of justice, to avoid intuitions that draw on the coercive nature of enforcing laws.*
justification is an account of the source of stability and part of what’s necessary for stability is individual citizens’ support for proposals. If we want principles and institutions to endure over time, then it is necessary for individuals to have reasons that apply to them and are compatible with their pursuit of the good life to act in accordance with those proposals. For clarity, and to avoid the complicated metaphysical question of what the fundamental nature of a reason is, we can have in mind something like Gaus’s explanation of what it means for an individual to have a sufficient reason. It says ‘Alf has (provisionally) a sufficient reason R if and only if a “respectable amount” of good reasoning by Alf would conclude that R is an undefeated reason (to act or believe)’ (2011, p. 250).

Second, public justification theorists endorse a sincerity requirement. Sincerity requires citizens to be able to sincerely offer other citizens reasons in favour of a proposal. The sincerity requirement imposes a different demand on citizens’ reasons than sufficiency. For citizens to be able to sincerely offer reasons to others, we need to think about whether the reason is sufficient for them in the way that it is for us. As Gaus says, ‘Justifying your beliefs or principles to others does not involve simply giving others reasons that they will accept, but in some way advancing reasons that you think are good reasons for them to accept’ (1996, p. 139). For a reason to ‘count’, then, it needs to be in some sense a reason for the citizen to whom it is offered and it needs to be in some sense intelligible to you. This feature requires a distinct consideration from that needed to establish sufficiency for yourself; you must consider the likely outcomes of others’ reasoning from their own perspectives. Moreover, for you to offer someone else a reason sincerely, it needs to be true that you, as the person offering the reason, also give it some kind of justificatory status. Without a principle of sincerity, the task of public justification falls into something more akin to persuasion or manipulation (ibid.). We could offer others whichever reasons we thought they would be persuaded by regardless of whether we took them to be good reasons. Publicly justifying proposals does not permit such actions and requires citizens to advance reasons that they ‘think are good reasons for [other citizens] to accept’ (ibid.).
3.3.2 The Convergence View

The distinction between consensus and convergence broadly captures the disagreement over how the public justification requirement ought to be satisfied. On the convergence view, a proposal is publicly justified when each citizen has sufficient reasons in favour of it. The content of the proposal is the only thing that citizens must agree over; their justifications for doing so may diverge. Let’s return to a topic where there are multiple divergent interpretations of the requirements of justice. If Jill and Jack constitute a society and they both think that abortion is morally objectionable under any circumstances, then that proposal would be publicly justified even if Jill’s reason was that the unborn foetus has an inalienable (secular) right to life and Jack’s was that the little bird on the windowsill told him so. A major implication of this view is that a proposal can be publicly justified—by each citizen having comprehensive reasons to support it—without being supportable by public reasons.

Jonathan Quong rejects the convergence view because it fails to ensure sincerity. Quong argues that we cannot sincerely expect others to endorse a political proposal when we can’t offer them public reasons in its favour because people can reasonably be expected to disagree over non-public reasons (2011, p. 256). Non-public reasons have comprehensive foundations and as such cannot justify a proposal to anyone who doesn’t share the comprehensive conception behind them. Public reasons, on the other hand, are based on a set of shared values that other members of the society accept. So, returning to our example, it’s implausible for Jill to think both that Jack’s comprehensive view is false and that it provides him with a good reason to prohibit abortion. If Jack lacks a good reason to support the proposal (from Jill’s perspective, since she doesn’t think listening to the bird on the windowsill is a valuable comprehensive doctrine), then she can’t sincerely offer him an alternative from her own, because she knows that he rejects it. Without shared reasons that are
sufficient for them both the sincerity requirement cannot be met (2011, p. 267).\footnote{Quong then concedes that if one believed something like Gaus's moderately relativist view about reasons and justification, then one could plausibly view many other people’s doctrines as false, but justified for those others to believe in. He then suggests that we can’t expect many reasonable people to agree on an epistemological theory like Gaus’s on account of the fact of reasonable pluralism (ibid., p. 269 ff.).}

Since sincerity is a plausible and important feature of an account of public justification, RU theorists must reject the convergence view in light of this objection. Like congruence, this view can only demonstrate stability in societies drastically different from those found in the world as we know it. The only circumstance in which Jill can sincerely offer a justification to Jack is if their comprehensive backgrounds are sufficiently shared for them to agree about the nature and priority of justice. RU theorists are not permitted to idealise the society to which their principles apply to establish such agreement (as we saw in the discussion of congruence).

Regardless of this objection, questions of stability are not pertinent in cases where pure convergence is possible. When each citizen has a comprehensive reason to favour a proposal, then it will be stable for the right reasons regardless of whether it is also supportable by public reasons. The real question for RU theorists is whether stability for the right reasons is possible given the fact of reasonable pluralism. As we have seen, this fact is a necessary feature of the moderately idealised world for which RU conceptions of justice are intended and so must be acknowledged for their conceptions of justice to meet the stability requirement. In order to show how the consensus view avoids this objection, I must introduce two concepts that feature in the account: the public political culture and a reasonable constituency.

3.3.3 The Public Political Culture

The importance of the sincerity requirement is clear once we remind ourselves of the purpose of public justification; to justify things to the reasonable public so that we can reasonably claim they can be motivated to act accordingly. The convergence view fails because comprehensive reasons cannot allow citizens to
sincerely offer reasons to others given the fact of reasonable pluralism. So, the public justification theorist needs to identify a set of reasons that allow citizens to meet the sincerity requirement in light of this fact. These are the shared, or public, reasons.

For many, an appeal to the Public Political Culture (PPC) (or something similar) is the way to do this. The PPC ‘comprises the political institutions of a constitutional regime and the public traditions of their interpretation (including those of the judiciary), as well as historic texts and documents that are common knowledge’ (Rawls, 2005, pp. 13-14). From these, public justification theorists establish a set of shared values that form the grounds of public reasons. ‘The political philosopher who pursues this objective [of stability for the right reasons] is looking for and/or striving to develop a shared fund of ideas that are latent in the culture of the community and that can be endorsed by people no matter what their metaphysical views’ (Hampton, 1987, p. 794).

For familiar societies, like modern Western democracies, we can draw on values that we know are widely accepted (for example freedom, equality, and fairness). Although not all citizens reason on the basis of these values in actual fact, they are the most fundamental substance that theorists can reasonably assume would inform citizens’ reasoning. Although I use the Rawlsian idea of a PPC, other public reason liberals ground the values on which public reasons are based similarly. Gaus says ‘unless our analysis of “true morality” connects up with what actual agents see as morality, our philosophical reflections will not address our pre-theoretical worries…our starting point must be what agents think they are doing when they judge and act morally—and our final destination cannot stray too terribly far from this’ (2011, p. 174). So, although Gaus isn’t deriving a set of shared values to secure sincerity, we can see a similar starting point in terms of the source of values theorists have to work with. Both reference actual agents in order to determine such values. Addressing the PPC ‘hopes to uncover, and help to articulate, a shared basis of consensus on a political conception of justice drawing upon citizens' fundamental intuitive ideas about their society and their place in it’ (Rawls, 1987, pp. 24-25).
The PPC is an appropriate source of shared values because the ideas within it ‘represent our most strongly held beliefs about justice. We take these convictions as provisional ‘fixed points’ in our political theorizing because we are most confident about their validity’ (Quong, 2011, p. 155). Pluralism amongst citizens and theorists mean that we lack any more fundamental basis for our claims without having to defend controversial conceptions of the good. The lack of consensus over the most fundamental and meta-ethical views of morality mean that such a defence is unavailable to RU theorists. Their purpose is not to discover such defences, but rather to provide action-guiding recommendations in light of these disagreements. If RU theorists took reasonable pluralism to be an insurmountable obstacle, then they would be unable to deliver on this aim.

3.3.4 Reasonable Constituency

Reasonable people—those who are moved to act according to terms that all citizens can accept (Rawls, 2005, p. 50)—are the appropriate constituency of public justification. This means that RU theorists need only justify their conception of justice to reasonable people in order to show its stability. A reasonable constituency is a particular idealisation of the citizens to whom conceptions of justice apply that is widely adopted by RU theorists in the literature. Nagel describes an idealised citizen as someone who ‘may be legitimately influenced in his conduct by these [personal] motives, but only in ways that he must acknowledge would be legitimate for anyone’ (1991, p. 31). Similarly, Scanlon responds to the idea that citizens must act according to terms that others can accept by appealing to ‘the idea of a shared willingness to modify our private demands in order to find a basis of justification that others also have reason to accept’ (1998, p. 5).

It’s perhaps odd that public justification applies to an idealised constituency like this rather than the entire public of actual citizens. This is perhaps odd because the intuitive appeal of public justification lies in the fact that, unlike congruence, it can respond to the fact of pluralism without making impermissible idealisations and the fact of pluralism is a very real issue that we face in the actual world. One might be concerned that justification to
reasonable people doesn’t address the difficulties of pluralism after all. Unreasonable people—those who fail to reason on the basis of society’s shared values—inevitably exist in any actual society. Public justification doesn’t ensure the support of unreasonable people and so we might question whether it is, in fact, a good indication that a conception of justice can be stable for the right reasons. We might think that the realistic nature of RU theorists’ societies should commit them to prescribing principles for a world that includes unreasonable people. However, since RU theorists’ constituency is not the actual world that we are currently in, some restriction of the constituency is permissible, and there is no prima facie reason, given their approach as we have considered it thus far, to suppose that they should or shouldn’t specify the restriction in this way. Theorists should, however, give reasons for whichever restrictions or idealisations they make. In this sub-section I consider reasons in favour of an idealised constituency of reasonable people and conclude that the idealisation is compatible with the RU approach.

Public justification that secures the support of unreasonable citizens is unnecessary to demonstrate stability for the right reasons. The essence of the idea of public justification, as Simmons puts it, is that ‘we want to take seriously the idea that we should offer justifications to people in their own terms without holding it hostage to the worst features of actual persons’ (1999, p. 763). Holding the idea of public justification hostage to persons’ worst features would force us to allow the content to be tainted (Bertram, 1997, p. 574), subject to errors in reasoning (Gaus, 1996, p. 121), and would make the account of stability political in the wrong way (Rawls, 2001, p. 189). The content of public justification is tainted, as Bertram puts it, when contingent factors, ideology, or the background effects of unjust institutions have undue influence (1997, p. 574). Bertram does not specify what he takes to be undue influence from such institutions, but a sensible suggestion is that influence is undue when citizens’ reasoning in favour of a policy or practice within an institution is not coherent or sound to anyone external to it. Theorists try to avoid this kind of bias by appealing to reasonable, rather than actual, people. I argued in the previous chapter that some contextual facts must play a role in the justification of some substantive principles of justice. However, this
idealisation to reasonable people is compatible with the requirement in the
previous chapter because my claim there included the proposal that principles
have some non-contextual grounds too. A constituency of reasonable people is
a way of permitting both contextual and independent normative considerations
to play a role.48 In places, I appeal to substantive commitments to ideals such
as freedom, equality, fairness, and respect that I assume all justice theorists are
in some sense committed to. These substantive assumptions demonstrate some
of the independent criteria by which we may judge the content of actual, rather
than reasonable, persons’ reasoning.

Focusing on the constituency of the reasonable avoids the difficulties raised by
pluralism because it allows theorists to focus on fair terms for each citizen as
free and equal despite their different conceptions of the good. Terms are fair
when they do not demand that each citizen conform to a particular ideal or
view of the good life, but instead propose a way for citizens to cooperate with
one another despite this. ‘To insist that everyone justify their free and willing
political commitments in the same way that we do is to insist that reasoning
must conform to a unique and preordained pattern; this is neither politics nor
philosophy, but simply fanaticism’ (Krasnoff, 1998, p. 290). When it comes to
public justification ‘all we can ask is that [citizens] freely and willingly comply
with the political claims that reason has justified’ (ibid.). Krasnoff’s point is
that individual’s reasons for supporting a proposal inevitably draw on their
conception of rationality, which is a part of their comprehensive conception.
Because theorists can’t assess and rely upon the details of each of those, they
may use the concept of reasonableness to show that the overlapping consensus
is available to each citizen in principle. Unreasonable citizens may not be
motivated to reason or act on the basis of such terms, but the terms’ status as
public means that they ought to do so. Public justification does not simply
impose a particular liberal conception of justice onto citizens because before
the justificatory process begins the content of publicly justified proposals are
indeterminate. It is their status as public (rather than liberal or something else

48 I say more about the relation to contextualism and the way in which the constituency of
reasonable people can reflect both contextual and non-contextual elements in the final section
of this chapter.
substantive) that makes these terms of cooperation fair for both reasonable and unreasonable citizens alike.

Understanding the idealisation to reasonable citizens in this way allows us to see how it is compatible with RU theories. It doesn’t make theorists susceptible to the guidance critique because it doesn’t test human motivations beyond a reasonable or feasible standard. The fact that some people are not motivated to act according to public beliefs is insufficient to say that it’s unreasonable to argue that people guided by a sense of justice are. Idealising a constituency to be made up of reasonable people is not only compatible with the RU approach, but it is in the spirit of the reason RU theorists use idealisations at all. This idealisation is helpful to RU theorists given that they are interested in the best foreseeable conditions and what humans could feasibly achieve in terms of justice (even if it isn’t very likely that they do). Whichever reasons theorists have for adopting the RU approach over alternatives, a commitment to a moderately idealised society with possible but favourable conditions is clearly compatible with a constituency of reasonable citizens.

This idealisation restricts the scope of individuals who must be able to support a conception of justice for it to be stable. Consensus theorists acknowledge this limited scope and shape the justificatory requirements necessary to achieve stability thusly. We shall see that this and the PPC are essential to this account’s success.

3.3.5 The Consensus View

Consensus theorists, unlike convergence, require proposals to be justifiable on the basis of public reasons that draw on values in the PPC. If a proposal is not justified by reasons sufficient for all reasonable citizens, then it is not publicly justified. The consensus view differs from convergence because it requires reasonable citizens to agree on both the content of the proposal and its justification. They think that ‘we can argue with one another about political issues in the name of our different visions of the human good while also recognizing that, when the moment comes for legally binding decisions, we must take our bearings from a common point of view’ (Larmore, 2002, p. 383).
Shared justification is only possible once a set of shared values like the PPC is established. In addition, consensus theorists claim that reasonable citizens must recognise shared values as the only appropriate source of public justification; any other reasons in favour of it do not contribute. Shared reasons have this special justificatory status in virtue of their PPC basis. Making shared reasons a requirement of public justification allows consensus theorists to avoid the criticism that a citizen’s endorsement of a proposal to her fellows will be insincere and based on reasoning they can reject. If Jill endorses a proposal on the basis of shared reasons drawn from the values in the PPC, then she can sincerely offer those reasons to Jack. Let’s say that the value of autonomous choice is prevalent in the PPC that Jack and Jill share. On this basis, Jill might (accurately, let’s say) reason that taxes on alcohol and cigarettes should not be permitted, because they undermine the autonomy of an individual’s choice to buy them. Jill is justified in thinking that such an argument will also be good reasoning for Jack, since he is a reasonable person and autonomous choice is a value in the PPC share. Whether or not Jack actually holds this view about the value of autonomy does not affect the sincerity of Jill’s actions. Its source in the PPC is sufficient to show that he should regard it as sufficient reason. Whether the reasoning itself is valid and sound—whether the claim that taxing cigarettes and alcohol undermines autonomy to a degree that we should avoid it is accurate—is not at issue. Jill’s conclusion on the basis of public reasons after a respectable amount of good reasoning is sufficient to establish sincerity when she offers it to Jack. The PPC, and the way that consensus theorists use values reflected in it, allows them to make a reasonable, rather than impermissible, assumption about the way that reasonable people are motivated to act. We can reflect this by re-stating the public justificatory principle. Consensus theorists might say something like ‘A coercive law L is justified in a public P if and only if each member i of P has sufficient reason R to endorse L, where R draws upon shared values.’ RU theorists may adopt this justificatory principle into their account, show their conception of justice to be stable, and avoid the objections to the alternative suggestions of congruence and convergence.
3.3.6 The Revised Consensus View

I want to propose a revision to the consensus view that is, as I see it, an improvement. The revision concerns consensus theorists’ commitment to permitting only shared reasons to contribute to the public justification of a proposal. Recall that citizens are reasonable in virtue of their recognising the justificatory weight of public reasons (‘those who are moved to act according to terms that all citizens can accept’ (Rawls, 2005, p. 50)), and so acknowledge that public reasons are necessary to meet the sincerity and sufficiency requirements of public justifiability. In this section I argue that a reasonable citizen also respects her fellow citizens as equals and thusly recognises their comprehensive reasons to have some justificatory weight even if they originate from a comprehensive background she rejects. In this way, comprehensive reasons can contribute to the justification of proposals in addition to the sufficient support they receive from shared reasons. I describe the revised consensus view in more detail, explain its relation to Rawls’s proviso (that non-public reasons can be given so long as shared ones are forthcoming) and other consensus theorists’ views, and then (in the following sub-section) demonstrate why RU theorists ought to prefer it to the conventional consensus view.

Shared reasons are necessary and sufficient for public justification on both revised and conventional consensus views. Distinctive to the revised view is the idea that additional justifications from comprehensive views are able to break ties and contribute to public justification. So, public reasons establish sufficient justification, but comprehensive ones can make a proposal more publicly justified. The revised view differs from convergence because it still requires public reasons for a proposal to be publicly justified, even if each citizen has comprehensive ones. It avoids the objection from sincerity in the same way as the conventional consensus view.

The revised view allows citizens to demonstrate a kind of respect towards one another that there is no room for in the conventional consensus view. In the event of a tie between two proposals that are both sufficiently publicly justifiable, the presence of comprehensive reasons can give citizens and theorists additional reasons to prefer one proposal over another. Moreover,
reasonable citizens should value these reasons regardless of whether they are rooted in their particular conception of the good. Their value is drawn from citizens’ respect for other citizens as deliberators. As we saw at the end of chapter two, equal respect requires that other citizen’s reasons in support of particular conceptions of justice be heard (Valentini, 2013, p. 195). Although public reasons are sufficient justification for each reasonable citizen by definition, the reasoning that motivates them and demonstrates their capacity for a sense of justice may be comprehensive and outweigh the public reasons from their individual perspective. As such, a role for these reasons is important, especially since the value of this kind of justification at all is its source in citizens’ own individual reasoning. For Vallier, this possibility is an essential aspect of an account of public justification. So much is clear when he says that ‘one crucial part of public reason that consensus theorists miss is that by restricting the set of reasons relevant to public justification, they lose the ideal of justification to the full reason of the individual, which is the only way to preserve her dignity as a free and rational being’ (2011, p. 273). Citizens should care about the source of others’ cognitive assent or/and their motivation, and caring about it means allowing it to contribute to public justification.

An example will help to clarify the approach. Jon, a third member of our hypothetical society, might think that abortion should be prohibited because the foetus has a God-given right to life. This is a sufficient reason to prohibit abortions for him, but he must recognise the necessity of public reasons for him to sincerely offer others. Since Jon is reasonable, he offers Jack and Jill public reasons to prohibit abortion on the basis of an ideal of autonomy for the unborn (since this value features in their PPC). Jon himself endorses the proposal because of the God-given right (a comprehensive reason) although the argument from autonomy (the public reason) is nevertheless sufficient. Distinctively, the revised account permits Jon to offer his comprehensive reason to Jack and Jill in addition to the public reasons. When he does this, the comprehensive reason contributes to the proposal’s public justification. Jack and Jill, since they respect Jon as an equal deliberator, understand the value of his being able to present his comprehensive reason in a public forum and
reason on the basis of it. The availability of deeper justification for those who share his comprehensive view is valuable to Jack and Jill, even if it doesn’t apply to them.

This idea is compatible with Solum’s ‘inclusive version’ of public justification that designates a ‘specified role’ for non-public reasons (1993, p. 748). That role allows the giving of non-public reasons iff it is the foundation for a public reason, or if it is an additional sufficient justification that can be given independent and sufficient justification by a public reason (ibid.). Rawls also embraces the idea that public justification can include a role for both public and comprehensive reasons. According to him, including non-public reasons into the public justification procedure is permitted, so long as the proviso is met. The proviso states that ‘reasonable comprehensive doctrines may be introduced in public political discussion at any time, provided that in due course proper political reasons—and not reasons given solely by comprehensive doctrines—are presented that are sufficient to support whatever the comprehensive doctrines introduced are said to support’ (1997, pp. 783-4). Weithman refers to this interpretation of public justification as the ‘wide view’ which ‘allows citizens to introduce and base their votes on comprehensive doctrine, provided that in due course they adopt and reason from that common viewpoint as well.” (2011, p. 330). Citizens can affirm the principle on the basis of their non-public reason, and have the public reasons in favour of it ‘ready’ for other citizens. It is not necessary for each individual to base her own support of the proposal on public reasons since, although sufficient, they may not be as weighty as her comprehensive reasons.

Weithman’s interpretation of the proviso is that citizens may ‘rely on their comprehensive doctrines…without adducing public reasons in support of their positions, so long as their doing so does not lead others to doubt that they acknowledge the authority of the public conception of justice. If doubts never arise, then the proviso is never triggered and they need do nothing more’ (ibid., p. 331). This helps to remind us that citizen’s support is the very aim of stability. Weithman thinks that this interpretation of public justification is the best view to take because it allows Rawls to solve the mutual assurance
problem which is, for him, the very reason that Rawls develops a conception of public reason in the first place. The mutual assurance problem is the idea that in order to secure stability, citizens need assurance that others will be guided by their sense of justice. Without such assurance, individuals might feel as though they are being taken advantage of (ibid., p. 49). For a conception of justice to be stable, it has to be true that citizens will support and uphold a principle on the assumption that others are likely to do the same. It’s likely that the revised view, which permits comprehensive reasons to have a role in the public justification structure, is likely to better solve the mutual assurance problem than the conventional consensus view can.

The main functional difference of the revised view is to act as a tie-breaker between cases in which two or more competing proposals are all publicly justified. In such cases, comprehensive reasons can contribute to the public justification of one proposal over the other by giving citizens respect-based reasons to prefer the option that receives the additional support. So, to return to our imaginary society, if there turn out to be autonomy-based public reasons to both prohibit and permit abortion, and one option receives additional support from comprehensive reasons, then all reasonable citizens should regard that option to be more publicly justified. Comprehensive reasons cannot make the difference between whether a proposal is at all publicly justified, but they can give additional justifications that reasonable citizens have reason to value.

In the following section I consider some of the reasons RU theorists should prefer the revised over conventional consensus view, but I don’t take these reasons to settle the case. They don’t settle it because the revised view depends on a particular conception of public justification with a controversial epistemic position that I haven’t sufficiently defended. The epistemic position says that someone else’s comprehensive reason can generate public support for a proposal, even when some members of the public don’t share its comprehensive foundations. Defending this position would require me to engage with epistemological questions of justification that are too tangential for this thesis.
3.3.7 Preferability of the Revised View over Conventional Consensus

I have said that the main functional difference between the revised and conventional view of consensus will be as a tie-breaker where two or more competing proposals receive equal support from public reasons. We can see from this role the first reason that RU theorists should prefer the revised view. Adopting it can improve the determinacy of the results of public justification. The results of public justification in the abortion example (where there were arguments both for and against) were, as we saw, indeterminate. RU theory doesn’t require each and every law to be determined at the abstract theoretical level, but requirements do need to be sufficiently determinate for theorists to test their ultimate implementability. I said in the previous chapter that permitting contextual facts to form a part of the grounds for some substantive principles ensures that RU principles are sufficiently determinate in this sense. Allowing particular comprehensive views to act as tie-breakers and inform the outcomes of public justification is a plausible way of meeting the contextualism requirement and could thus be a reason to support the revised over conventional consensus view. I say more about this possibility in the following section of the chapter.

A second reason that we might prefer the revised view is because public support will be more stable when comprehensive reasons can enter the public forum and count in favour of a proposal’s public justification. RU theorists don’t require more stability, since shared reasons are sufficient to meet the stability requirement, but it’s plausible that they should prefer it. Although sufficiency is good enough, it doesn’t mean that there isn’t something better when it comes to public justification. This is especially the case when we consider the tie-breaking role I have imagined for this approach. Let’s return to the example in which plausible arguments from shared public reasons could be made both in favour of and opposing the proposal that abortion should be prohibited. If only one of these options received additional comprehensive support, as we supposed in the example, then we can make a stronger case for the stability of that proposal than we can for the alternative that only receives public support.
Finally, if citizens may offer comprehensive reasons in support of proposals, then the sufficiency requirement is met in a more authentic way. The sufficiency requirement is fundamental to public justification and so we should prefer the view that best permits us to satisfy it. The revised view does this because it permits proposals to draw support from congruent conceptions of the good (where such congruence exists). Recall that congruence was rejected as a requirement of stability not because it fails to secure it, but because it isn’t feasible (even on a very thin account of feasibility). Allowing comprehensive reasons to contribute to public justification is a more authentic reflection of citizens’ capacity to reason. As a reasonable citizen Jon can recognise the value of autonomy for the unborn (found in the PPC) to be sufficient to explain why abortion shouldn’t be practised in his society. But in making this explanation he might feel as though he is being in some sense deceitful if he cannot also explain that the source of his assent is in-fact the God-given right to life. Other citizens’ rationality is valuable and by allowing non-public reasons to play this kind of role, and permitting citizens to ‘appeal to some of their most compelling commitments’ (Vallier, 2011, p. 264), we can recognise them as such. What motivates a citizen to support a proposal is her interpretation of what she thinks justice requires given her conception of the good, and oughtn’t to be omitted from public deliberation. Solum’s reasoning in favour of his ‘inclusive view’ is partly informed by a reason like this when he says that the disclosure of non-public reasons reinforces authenticity (1993, p. 751). To ask some citizens to refrain from offering what they see as the true grounds for a proposal undermines their sense of justice and makes the justification less ‘public’.

So, theorists may prefer the revised view of public justification because doing so makes public justification as a means to stability more determinate, stable, and authentic. Although these are good reasons, they are insufficient to make the case conclusively in favour of the revised view. Since theorists must adopt some account of public justification, and the conventional consensus view is sufficient to meet the sufficiency and sincerity requirements, it too is acceptable for RU theorists to meet the stability requirement.
3.4 People’s Views and the Stability Requirement

At the end of the previous chapter I tentatively introduced the suggestion that people’s views about justice were the appropriate contextual facts theorists might consider in order to meet the weak contextualism requirement. I said that doing so was necessary to treat citizens as equal individuals, but did not conclusively defend this argument. In this section, I outline the way in which sensitivity to people’s views is also compatible with the demands of the stability requirement. People’s views about justice can inform the stability of a conception (via the PPC) and simultaneously meet the criteria of contextual facts to which RU conceptions must be weakly sensitive. In this sense this idea can unify the two methodological requirements. People’s views inform stability despite the fact that the appropriate constituency of public justification (i.e. those citizens to whom proposals must be justified) is reasonable, rather than actual, citizens. I demonstrate the role people’s views have informing stability and then discuss the compatibility of this conclusion with the requirements of weak contextualism. As we shall see, the nature of the PPC is such that fulfilling it will permit theorists to meet the weak contextualism requirement simultaneously regardless of its particular conception. Since RU theorists should accept the necessity to meet the stability requirement, this gives us another reason to accept the idea that contextual facts will form a part of some principle’s grounds.

3.4.1 Actual People’s Views and the Public Political Culture

People’s views about justice inform stability because of their role in the PPC. The PPC is the basis of shared values that proponents of both the consensus and revised views endorse as the appropriate source of justificatory reasons in the process of public justification. As we saw, public reasons draw on the PPC because it represents the most settled convictions of a society. For a theorist to demonstrate her conception of justice to be stable for the right reasons, she must show it to be consistent with these values. Their shared nature means that theorists can permissibly assume that reasonable citizens will be motivated to reason on their basis. To expect citizens to reason on the basis of any other values is to deny the nature of their disagreement with one another as
reasonable. Since the PPC is valuable because of its shared nature, it must in some sense reflect the views of that society’s members (otherwise it wouldn’t really be shared). The status of the values in the PPC (as those that proposals must draw on in order to be stable) is established by their source in the views of citizens and it is via this function that people’s views have an important role in the RU theory of justice. It’s important to note that the restriction to reasonable people is an idealisation of the constituency to which conceptions of justice are to apply, and does not similarly restrict the constituency that informs the PPC. Reasonable people are defined as those who are motivated to act according to values reflected in the PPC; it does not necessarily follow that they are the exclusive source of such values.

People’s views might inform the PPC in numerous ways, depending on the particular conception we favour. Consider the following example. Historically, the members of a society support a value but eventually the status of that value comes to be eschewed and it then no-longer features in the PPC of that society. For example, duelling was thought to be morally permissible in eighteenth century Great Britain because the value of honour was so important that it sometimes warranted harming others (Appiah, 2011). Eventually, by the middle of the nineteenth century, the British public had changed their view and honour no-longer held the justificatory status to permit duels (ibid.). The change in the PPC is the direct result of citizens’ attitudes regarding the value in question. What shapes people’s attitudes is a complicated and disputable matter, but we can be sure that such views are a good measure of the prevalent values in society at any one time since they can reflect the culmination of many other factors. These other factors (the political institutions and the public traditions of their interpretation, historic texts and documents that are common knowledge) may feature in the PPC alongside people’s views as well as by shaping their content. People’s views are necessarily included in this list because of the PPC’s purpose to reflect the shared values of a society. This influence of people’s views and attitudes is true of many interpretations of the PPC whether theorists take it to be: an implicit or explicit function of the majority’s views (via a voting system); an overall set of values that makes the disparate views as consistent as possible (provided by social scientists); a model
that aggregates the preferences of everyone in a society as best as possible (another task for social scientists); or a set of values that haven’t been vetoed by any groups in the society (again via a kind of voting system).

Theorists may wish to impose constraints onto the permissible content of a PPC. A theorist might, for instance, suggest that the view expressed in some cultures that baby boys are more valuable than baby girls cannot feature in the PPC of any plausible account of justice. As such, she might impose constraints to prevent such objectionable values from being included. The presence of constraints does not prevent people’s views from informing the stability of a conception of justice entirely because it does not mean that their influence is wholly withdrawn. Their views are still a necessary input to the PPC. As I argue in the previous chapter, universal deontological constraints are not determinate and so their presence does not mean that the set of values in the PPC is fixed. As such, people’s views (and other inputs to the PPC) can still contribute and their input won’t always be over-ruled by the constraints. The existence of these constraints doesn’t mean that we abandon the idea that shared values found in the PPC are at all valuable.

Even given these kinds of constraints, some theorists may resist my interpretation of the PPC and its mildly relativistic nature. There are several examples in history where the prevailing view of a society is not something that most analytic political theorists would think could inform any plausible conception of justice. Historical attitudes to slavery in the US are an obvious such example. There are three reasons that theorists should not be troubled by such concerns. First, as I outline in the previous paragraph, RU theorists can rule out some views by adopting a set of deontological constraints. Second, the problem with such views is often not the values that proponents appeal to, but rather empirical beliefs that we know to be false (Miller, 2002, p. 23) encouraging people to misapply their values. Supporters of slavery weren’t necessarily rejecting the value of equality, but rather misapplied it in the case of

\[49\] Deontological constraints would, of course, require justification. Justifying such things is a task perhaps best pursued by adopting the ToI approach to justice, the justificatory demands of which I leave open.
slavery on the basis of the (obviously mistaken) empirical idea that black people were somehow different from white people in a way that warranted different treatment (for instance on the basis of phrenology). Third, the concerning aspect of views like this will often be overruled if we consider the views of all members of society or sides of historical/cultural traditions. If the PPC represents all citizens’ views (including the slaves themselves) then the troubling prevailing view may not be as dominant as we expect.

As I have tried to show, there is a role for people’s views insofar as they inform the PPC. Clearly, the extent of their influence depends on the particular conception of the PPC and the way in which various inputs shape the values we draw on from it. The stringency of the deontological constraints, if there are any, will also bear on the influence of people’s views. If we accept the revised consensus view of public justification, then citizen’s views have a more significant role since individual comprehensive views can have a role as well as those informed by the values in the PPC. When comprehensive views contribute to the public justification of a proposal, the role for individuals’ views is much clearer.

So, people’s views inform stability via public justification in two ways. Actual people’s views inform the PPC, and thus inform the values that public reasons are based upon. Furthermore, reasonable people with comprehensive reasons in favour of publicly justified proposals can influence stability because the revised view allows such reasons to contribute to a proposal’s public justification.

3.4.2 Contextualism and the Public Political Culture
In the previous chapter I defended the weak contextualism requirement. That is, I argued that the grounds of some of substantive RU principles must be partially given by the context to which those principles are to apply. Contextual facts are non-universal and reflective of the shared characteristics of particular communities. When contextual facts form part of the grounds of some substantive principles of justice, then the theory is sensitive to them and weakly contextualist in the appropriate sense. At the end of the chapter, I suggested
that people’s views may be a good source of a contextualist fact (without ruling out the possibility of other types of contextual facts playing a role too).

If people’s views inform the PPC in the way that I have suggested in the previous subsection, then RU theorists may meet their commitments to stability and contextualism simultaneously by including one justificatory principle that makes substantive principles sensitive to people’s views about justice. The values found in the PPC are contextual since they are not universally shared and, as such, they reflect the shared circumstances of fellow members of the society. Requiring a conception of justice be sensitive to them ensures that the conception of justice is sufficiently determinate (to meet the contextualism requirement), and that reasonable citizens can be motivated to comply with its requirements (to meet the stability requirement). When theorists draw on the values in the PPC to demonstrate a conception’s stability, those values form a part of the conception’s grounds. In this way, actual people’s views have a role in the justification of stable conceptions of justice.

Moreover, RU theories are necessarily sensitive to people’s views since they must feature in any plausible conception of the PPC, and the shared basis of the PPC is essential to the success of the consensus views of public justification. This does not mean that other contextual facts cannot also feature as a part of RU principles’ grounds. I revisit the proposal that principles of justice ought to be sensitive to people’s views once again at the end of the following chapter, to see whether this unifying theme extends to the third and final methodological requirement I defend.

Conclusion

In this chapter I have shown that RU theorists ought to accept the stability requirement and show their conceptions of justice to be capable of ordering society over time. They ought to do this because it’s implausible that feasibly implementing a conception is relevant to the content of justice, but that the stability of the institutions that support it is not. Moreover, theorists ought to demand stability for the right reasons because the alternatives either conflict
with their substantive commitments (to freedom and equality) or fail to recognise that conceptions of justice must give citizens reasons to act from their sense of justice.

RU theorists must secure this kind of stability via public justification because the alternative, congruence, requires them to make impermissible idealisations. Although both the consensus and revised views of public justification are plausible, and compatible with the RU approach, the latter is preferable for several reasons. The requirement that RU conceptions of justice must be the subject of public justification is somewhat demanding and narrows the range of permissible methodologies significantly. Although I haven’t specified a precise justificatory principle, there’s only two plausible views that I discuss (consensus and revised). It could be the case that others are also compatible although it’s unclear what form they would take so that they could meet the requirements of public justification and avoid the criticisms levelled at the alternatives discussed.

At the end of the chapter, we saw that fulfilling the stability requirement in this way involves a role for actual people’s views informing the content of justice. This role is necessary because the shared status of the PPC means that people’s views are an essential input. A PPC, or something similar, will feature in any plausible account of public justification that fulfils the two demands I identify. This conclusion draws support for the tentative conclusion of the previous chapter and unifies the two requirements established so far.

Although stability ensures that reasonable citizens will be suitably motivated to adhere to the requirements of a particular conception of justice, implementing it will still necessarily be coercive. The coercive element is necessary because even reasonable citizens need assurances that others will comply and the state provides such assurances via the use of coercive threats for non-compliance. In the next chapter I consider the circumstances in which such implementation is permissible.
4. Legitimacy

Introduction

We have seen so far that contextual facts and stability should form a part of the grounds of RU conceptions of justice in addition to the fundamental requirement that principles be feasible. Each of these commitments results from a claim about the nature of an RU conception of justice and what’s necessary to justify it. In this chapter, I argue that RU conceptions of justice ought to be legitimately implementable (this is the legitimacy requirement). For a conception of justice to be legitimately implementable, it needs to meet certain procedural or substantive requirements (given by an account of legitimacy) that explain the circumstances in which someone could permissibly use coercion to impose it. As such, RU theorists should not recommend any laws/policies/actions that are illegitimate. When a particular conception of justice meets these requirements in a particular time and place, then its imposition is legitimate in those circumstances.

To consider political legitimacy is to consider the empirical fact that the state coerces citizens by imposing its will onto them and forcing them to comply. We normally assume that citizens should be free to lead their lives as they wish, and so if the state forcing them into particular actions is permissible, then we naturally ask what gave it the right to do so. The literature surrounding this question is vast and tightly linked to the corresponding question of whether citizens have a moral duty to obey. In this chapter I am interested in the
coercive nature of implementing conceptions of justice and the possibility that a legitimacy account can successfully explain the circumstances in which coercion is compatible with justice. As such, I am not concerned with the source of political obligation. I don’t defend a particular account of legitimacy, but rather the idea that RU theorists must adopt a plausible account to test whether their conception is legitimately implementable.

Imagine that an RU theorist has a provisional conception of justice that she takes to meet the justificatory requirements I have established so far and whichever other demands she adopts (recall that a range of demands are possible within the RU approach, albeit limited by the necessary inclusion of those I’ve defended). The theorist must then test her conception of justice to see whether it is legitimately implementable. The necessity of this test imposes an additional methodological commitment on the approach. The theorist selects her preferred account of legitimacy—which establishes constraints on the content of conceptions of justice—and then judges whether the requirements of her conception fall within these constraints. If the conception of justice passes the test, then it is legitimately implementable. Legitimate implementability contributes to the grounds of a conception of justice in the same way that stability and contextual facts are also necessary conditions of justification.

In section one I defend the legitimacy requirement. RU theorists should demand that their conceptions of justice are legitimately implementable because the presence of illegitimate coercion is morally problematic. Section two is spent considering the most prevalent accounts of legitimacy and whether they are suitable for theorists to adopt them to test the legitimately implementability of their conceptions of justice. I illustrate this process by appealing to the familiar substantive conception of Justice as Fairness (Rawls, 1999) and working through the legitimacy ‘test’ in the way described above. I consider two versions of the natural duty of justice, the natural duty of rescue, consent theory, fair-play theory, and democratic legitimacy. Consent, fair play, and a particular version of the natural duty of justice are incompatible with the methodological requirements of the RU approach established thus far. Finally,
in section three, I consider whether the legitimacy requirement draws further support for the suggestion that people’s views are in some way relevant to the content of RU theories of justice.

4.1 Why Legitimately Implementable?

I understand legitimacy to refer to the conditions in which individuals or groups have the moral right to coerce others into following their directives. Understanding legitimacy in this way distinguishes this right to coerce from the moral right to impose duties onto others (this is what I understand by authority). Estlund also adopts this distinction and expresses it in the following way: ‘the authority of a command—its power to create a requirement to obey—is conceptually separate from its legitimacy (the question of whether coercive enforcement would be permissible)’ (2008, p. 42). These definitions of the concept of legitimacy are widely disputed. I don’t conclusively defend this conception over alternatives, but this view’s central feature of legitimacy’s coercive nature makes it particularly relevant for RU theorists. RU theorists should not endorse a conception of justice that requires illegitimate coercion, and so its content is constrained by the requirements of a legitimacy account. Whether or not the moral permissibility of coercion is properly labelled as a legitimacy concern, this is the important and relevant consideration for RU theorists.

It’s important to consider the source of legitimacy because coercing an individual is prima facie wrongful. Its wrongful nature may be grounded in an ideal of freedom (an individual is unfree if she is forced to perform or avoid specific actions (Raz, 1986, p. 410)) or autonomy (coercion subjects an individual to the will of another and thus violates her independence (ibid., p. 378)). Regardless of the exact nature of the concern, I presume that RU theorists are committed to the idea that individuals have some kind of right to self-govern, and that this makes at least some instances of coercion of one individual (or group) by another (individual or group) morally problematic. Imagine Jenny is in a shop and wants to buy chewing gum. Johnny comes along and demands that she buys vimto instead. He says that if she doesn’t, he
will forcibly punish her (he says that he will lock her in the shop for 24 hours). In this case Jenny was coerced into buying vimto and we would usually think that this meant she had been wronged. If Johnny’s actions were legitimate, then his action would be sufficiently justified and we need no longer consider them wrongful.

Legitimacy and authority are particularly important in politics because of the nature of the state. States are coercive when they force citizens to obey the law (directly or via threats) or when they issue punishments for those who disobey. Political institutions like the state are just when they fully respect their subjects’ rights (Valentini, 2012, p. 595), including the right to self-govern. Political legitimacy requires us to consider the circumstances in which the state can permissibly coerce citizens in these ways. In a moderately idealised utopia the laws (and thus the state’s coercive actions) are informed by a conception of justice. The prima facie wrongfulness of coercion is problematic for a conception of justice; we can’t say that Johnny was acting justly in the shop unless we explain the source of his right to coerce Jenny. Theorists appeal to an account of legitimacy to make such an explanation. Perhaps Jenny’s blood sugar was dangerously low and Johnny prevented her from harm by forcing her to have the sugary drink. If prevention of harm to oneself is a legitimate justification for coercion, then Johnny acted justly. Similarly in the political context, an account of legitimacy explains in what circumstances the state has the right to coerce citizens. These accounts impose demands on the nature of the state’s coercive actions (and thus conceptions of justice). If these explanations are successful, then the prima facie wrongful act of coercion is just and the conception of justice is legitimately implementable.

In the first chapter I discussed the possibility of political facts, such as the coercive nature of the state, as something which feasibility concerns may make RU conceptions of justice sensitive to. As we saw, the presence of political coercion, or any other kind of political fact, is not something that is necessarily

---

50 I assume throughout the chapter that the actions of the state are given (or at least constrained) by a conception of justice and that it is their role to order society according to a conception in this way.
true of all possible worlds and so is not something that imposes a feasibility constraint (according to the uncontroversial conception). Rather, legitimacy constrains justice because RU conceptions of justice must be ultimately implementable into a version of the world as we know it. This makes them sensitive to the prima facie wrongful nature of coercion because such a nature is problematic for a conception of justice and must be explained (rather than because legitimate coercive implementation is infeasible). The wrongful nature of coercion means that it is impermissible for the state to coercively impose an otherwise just conception of justice without an explanation from an account of legitimacy. Doing so would violate citizens’ rights to self-govern (or whichever right the prima facie wrongfulness of coercion is grounded in). The following example clarifies this role.

An RU theorist might think that some form of luck egalitarianism is the most plausible conception of justice, and find justifications to support her view (let’s say that contextual facts contribute to its grounds and that it can be the subject of public justification and thus stable for the right reasons). However, she soon discovers that the state cannot legitimately redistribute in order to neutralize the effects of bad brute luck as the conception requires. The theorist has tested her conception of justice against the account(s) of legitimacy that she finds plausible (and that are otherwise compatible with her approach) and found that there’s no justification for the state to coerce citizens into meeting the demands of luck egalitarianism. Since RU theorists must ensure their conceptions of justice are legitimately implementable, she must reject or revise luck egalitarianism in light of its failure to pass the legitimacy test.

So, a conception of justice is legitimately implementable if its recommendations are within the constraints set by legitimacy. If legitimacy requires democratic support, then a conception of justice is only legitimately implementable if it can be shown to receive such support in the society for which the RU theorist is designing principles. If all the recommendations of a conception of justice are within these constraints, then that conception is legitimately implementable. In this way, legitimacy informs the content of justice. Once a conception is shown to be legitimately implementable this fact
forms a part of its grounds. Whether or not the constraints imposed by legitimacy are substantial depend on the particular account. The luck egalitarian example demonstrates how this could happen, but common understandings of legitimacy reveal that we should not expect these constraints to be particularly narrow. If the legitimacy account sets broad constraints then actions could be legitimate without being just (formal equality of opportunity might be an example of this). For Rawls, ‘legitimacy is weaker than justice and imposes fewer constraints on what can be done’ (2005, p. 428). Peter says that legitimacy is, although a weaker concept than justice, ‘in some sense, more fundamental...In the Rawlsian view, moral justification thus incorporates the concept of legitimacy; the two are not independent ideas. Thus prioritising legitimacy also has the practical implication that the only way to achieve justice is through a democratic process that satisfies the conditions of legitimacy’ (Peter, 2009, p. 62). Even if legitimacy does not make a substantive difference to the content of justice, a conception’s status as legitimately implementable contributes to the justification of the substantive conception of justice. Legitimate implementability is necessary, but not sufficient, for justice. I now elaborate further on the reasoning as to why legitimacy has this role in an RU theory.

This understanding of the relation of legitimacy to justice requires theorists to reject the view that they are entirely separate concepts. It might at first seem plausible to say that legitimacy is a concern completely separate from the content of a theory of justice. If this were the case, then legitimacy would have no place in this thesis and the claim in this chapter would be misguided.\footnote{It is worth noting that this presumption conflicts with an alternative idea: that the content of legitimacy is entirely given by justice so that an account of the former can never affect the content of a conception of justice. I consider (and reject) this account of legitimacy in the following section.} Consider theorists who adopt the Theory of Ideals approach. These theorists aim to develop an account of justice that provides an accurate reflection of the truth of the ideal. For them, the fact that coercive force is necessary to implement the requirements of a conception of justice, and the fact that such coercion is prima facie wrongful, would fail to undermine the justice of their
ideal. It’s not intuitively obvious that such an empirical fact necessarily shapes the content of the ideal of justice, and so their methodology does not require them to consider it. RU theorists, on the other hand, are fundamentally committed to producing a conception of justice that guides the constituents of a moderately idealised utopia. This means, as I outlined in the first chapter, that their conception of justice must be ultimately implementable (this is what distinguishes their approach from the Utopian alternative). This means that although their conceptions of justice are designed to govern the constituents of a utopia, theorists are constrained to adopting substantive principles that could be implemented into some possible version of the world as we know it. The permanence and nature of political power in the world as we know it mean that the prima facie wrongfulness of coercion is something that always constrains RU conceptions of justice and requires theorists to adopt an account of legitimacy. Once the requirements of their conception of justice are shown to be within the constraints imposed by the legitimacy account, its legitimate implementability may contribute to its grounds.

First, the permanence of political power refers to the fact that an organising body like the state is, and always will be, necessary to make sure that citizens benefit from co-operating with one another. As we saw from the fact of pluralism in the previous chapter, people have conflicting ideas about justice and the good life and so their ends (and the ends they see fit for justice) differ. This conflict creates the need for an organising body to arrange affairs so that citizens pursue the same public goals and don’t violate one another’s rights while doing so. ‘Only an agency vested with an actual monopoly of force has any hope of being able to act in a coordinated fashion on that sort of scale…If different people use force to secure ends which contradict one another, then this systematicity is lost’ (Waldron, 2006, p. 194). Moreover, even if everyone shared the same ends, the coordinating role of the state would still be crucial to ensure that people pursue these shared ends in an efficient and peaceful way.

In addition to its permanence, political power is always a coercive power applied by the state (Rawls, 2001, p. 40). The idea that it is essentially coercive reinforces the RU theorist’s need to adopt an account of legitimacy in order to
reconcile her conception of justice with the prima facie wrongfulness of its implementation. The very nature of a state incorporates the idea that it has a monopoly on the permissible use of force. The use of such force is usually necessary to coerce non-law-abiding citizens. However, as we saw in the previous chapter, RU theorists may idealise the constituency to which their principles apply. If they idealise citizens to be reasonable or something similar, then we might think that non-law-abiders are not a problem for their approach. If the conception of justice is stable, then the theorists will have shown that citizens can be motivated to comply with the same conception of justice despite pluralism. In this case, the state may never need to force citizens into compliance because stability shows that they’re motivated to do so. These suppositions about the implication of an idealised constituency would, however, be mistaken. Despite the stability of a conception of justice, coercion is a necessary presence in any state-governed constituency because even reasonable citizens need assurances that others will comply. ‘The need for enforcement of rules by the state will still exist even when everyone is moved by the same sense of justice’ (Rawls, 1999, p. 236). The threat of force gives citizens the assurances they need and so ‘some device for administering fines and penalties must be established’ (ibid., p. 238). Any stronger idealisation to remove this feature of citizens will result in an infeasible conception of justice. Recall that feasibility requires success to be reasonably likely conditional upon trying. ‘In a large community the degree of mutual confidence in one another’s integrity that renders enforcement superfluous is not to be expected.’ (ibid., p. 237) Without such assurances, a conception of justice could successfully order society. In this sense political power is always coercive even when a society is ordered according to a conception of justice that is stable for the right reasons. The necessity of the threat of force to reassure citizens of others compliance shows why political power is always coercive, even in idealised constituencies.

Moreover, idealising political power to suppose that it is not coercive is incompatible with Valentini’s limitation on idealisations. Recall that this limitation prohibits idealisations that presuppose just background conditions of the circumstance that the principles are trying to correct (2009, p. 352). Idealising political power so that it is non-coercive does exactly this when we
understand the prima facie wrongful nature of coercion to be problematic when implementing a conception of justice. The purpose of the concept of legitimacy is to explain the circumstances in which such exercises of power are permissible. So, if theorists idealise the nature of political power to be non-coercive, then they assume away the problem that accounts of legitimacy are supposed to correct. The ultimate implementability requirement of the RU approach to justice demands that theorists acknowledge the political facts of a permanent coercive state as conditions that will always face the world as we know it. Acknowledging these facts requires them to recognise the prima facie wrongfulness of coercion and explain how it can be compatible with the just implementation of their substantive principles. Conceptions are legitimately implementable when their contents are compatible with this explanation.

So, the requirement for conceptions of justice to be legitimately implementable is not the result of illegitimate conceptions being infeasible, but rather the only way for theorists to respond to the demand for justification created by the prima facie wrongful nature of the inevitable coercion imposed by a political power. That justification comes in the form of an account of legitimacy. Without a legitimate explanation, the presence of prima facie wrongful coercion undermines the justifiability of the substantive principles of justice. For this reason, RU theorists cannot develop a conception of justice that is blind to considerations of legitimacy and must instead acknowledge that, for them, part of what makes a conception just is the fact that it is legitimately implementable. This makes the effects of the legitimacy requirement similar to those we saw in previous chapters. Like the others, this requirement imposes a particular justificatory demand that narrows the range of methodologies and resulting substantive principles available to RU theorists.

I now consider some of the many competing accounts of legitimacy. These accounts specify what reasons and values or what kind of a conception of justice can explain legitimacy (Rawls, 2001, p. 41). Each identifies particular characteristics necessary for a coercive act to be legitimate—they might say that the act must be a requirement of justice, or result from a democratic procedure for example. When an RU theorist considers whether the state
could legitimately implement her conception of justice, she must test whether its requirements are compatible with the characteristics specified by the legitimacy account. I run this test using the familiar substantive principles of justice as fairness as a conception to test against various existing accounts of legitimacy.

4.2 Existing Legitimacy Accounts

In order to meet the legitimacy requirement, the RU theorist must adopt an account of legitimacy and show that implementing the requirements of her conception of justice is or could be within the constraints of legitimate coercive action it establishes. Doing this demonstrates that her conception of justice is legitimately implementable. As I have said, an account of political legitimacy is a particular explanation of the source of the state’s right to coerce citizens that explains how its coercive implementation of a conception is compatible with justice. It’s important for me to test whether existing accounts of legitimacy are plausible given the justificatory role legitimacy is to have in RU theories. It must be at least in principle possible for a conception of justice to fulfil each RU justificatory requirement simultaneously. For example, an account of legitimacy that permitted coercive action on the basis of universally grounded principles exclusively could never be compatible with the requirement that contextual facts form a part of some RU principles’ grounds. If the legitimacy grounds are in principle compatible with stability, feasibility, and contextualism requirements, but clash when it comes to a particular conception of justice, then the theorist can decide a trade-off to determine which she takes to be the weightiest. So, if the actions required to implement a substantive libertarian conception of justice were all legitimate but not supportable by public reasons in a particular context, then the theorist could choose whether to argue that the libertarian conception was preferable to illegitimate alternatives supported by public reasons. Although she can argue in this way to establish preferability of available options, if a conception of justice is not both stable and legitimate, then it cannot be a fully justified conception on the RU view since both are necessary conditions.
In this section I take Rawls’s (1999) justice as fairness (JAF) and imagine myself to be an RU theorist testing this substantive conception of justice to see whether it is legitimately implementable. JAF famously contains the following two substantive principles:

a) Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all; and

b) Social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they are to be to the greatest benefit of the least-advantaged members of society (the difference principle).

(Rawls, 2001, pp. 42-3)

I am presuming for the purposes of this test that JAF meets the other justificatory demands of an RU theory. That is to say that its implementation is feasible; that contextual facts form a part of its grounds; and that its content is supportable by public reason (and thus it is stable for the right reasons). JAF also meets several other justificatory requirements, as per Rawls’s particular methodology. These additional grounds vary for different RU theorists; they may adopt alternative methodologies within the constraints of the approach as I have established them. Since we are imagining that JAF is at the stage of theory construction, its contents are provisional. We must test them against the requirements of particular accounts of legitimacy to see whether this conception can meet the justificatory demand imposed by the legitimacy requirement.

Political legitimacy is one of the oldest and most disputed issues in political theory. For this reason I restrict my focus in this section to what I take to be the most influential accounts. I first consider accounts of legitimacy that specify the source of the state’s right to follow from citizens fulfilling a natural duty. The natural duty of justice says that coercive control by the state is permissible when it is necessary for citizens to best fulfil their duty of justice.
This imposes a substantive requirement on a conception that its implementation must secure justice in order to be legitimate. The second natural duty account that I consider is the duty of rescue. This account explains the state’s right to coercive rule as a liberty right possessed in virtue of citizens’ lack of right to be free from interference. Citizens lack this right because of a duty of rescue owed to other citizens. This account imposes the substantive requirement that a conception’s recommendations be necessary to facilitate citizens fulfilling this duty of rescue. Following this, I consider the most prevalent account of legitimacy that relies on an acquired right: consent theory. According to this account, a conception of justice is legitimately implementable when the state has acquired the right to use coercive force to implement it via citizens’ permission. The second account of legitimacy as an acquired right is fair play theory. On this view citizens must obey the state to fulfil a duty of fairness they owe to their fellow citizens. The enforceability of this duty depends on its fulfilment being necessary to avoid violating the harm principle. A conception of justice must be similarly necessary for it to be legitimately implementable. Finally, I consider democratic procedures as a source of legitimacy. On these views, legitimacy does not issue a substantive demand based on the requirements of citizens fulfilling a particular duty, but rather justifies the state’s coercive actions on the basis of their being the result of a particular procedure. In this case, a conception of justice may be legitimately implemented if it can be the result of such a democratic procedure.

As we shall see, several of these leading accounts are compatible with the justificatory role for legitimacy and the other commitments of the RU approach. I take this fact to support the idea that legitimate implementability contributes to the grounds of a conception of justice. I do not defend any legitimacy account as attractive or plausible outside of its compatibility with the other requirements of the RU approach. Each is briefly outlined via a leading proponent of the view in order to consider whether it can perform the justificatory role that I attribute to legitimacy here and is compatible with the methodological commitments of the RU approach already established. It is up to particular theorists to decide which account to adopt into their theory.
4.2.1 Natural Duties

Natural duties are moral obligations that we all owe without performing any specific actions beforehand. We possess them in virtue of our status as autonomous beings. The moral obligation not to harm others is an example of a natural duty that we all possess regardless of our particular circumstances. In this section I consider two distinct natural duties, each of which is said to correlate with the state’s right to the legitimate use of coercive political power. If the state coercively imposing JAF is a means to citizens fulfilling these duties, then it is legitimately implementable. I consider the natural duty of justice first.

4.2.1.1 Natural Duty of Justice

A dominant view in contemporary political theory is that showing a state is justified and showing it is legitimate require the very same arguments (Simmons, 1999, p. 739). The natural duty of justice account stipulates that a set of rules can be coercively enforced so long as those rules are requirements of justice. Since each citizen has a natural duty to do as justice requires, coercive actions that force her to fulfil these requirements are permissible and unproblematic for a conception of justice. The view’s fundamental claim is aptly described by Quong: ‘one way to establish that a person has legitimate authority over another person involves showing that the alleged subject is likely better to fulfil the duties of justice he is under if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to directly fulfil the duties he is under himself’ (Quong, 2011, p. 128). I describe what needs to be true for a conception of justice to be

52 An attraction of Quong’s view is that it draws on the appeal of Raz’s (1986) famous normal justification thesis. Raz’s view grounds legitimate authority in practical reason so that submitting to the coercion of another is permissible when doping so allows you to better comply with reasons that apply to you (ibid., p. 53). Quong takes this account and shows how it can be adapted to yield the good of justice, rather than practical reasoning more generally. Raz’s account is omitted from more detailed discussion here because its purpose is not to establish an enforceable right held by the state but rather the conditions under which an individual has the moral duty to submit her will and judgement to those of another (Raz, 2006, p. 1012). Of course one could take the relation between citizen and state and try to demonstrate such conditions exist and the account can explain political legitimacy. An RU theorist would then need to show that each citizen has reasons to comply with the requirements of her conception of justice and that the duty to comply with them was enforceable. Only then would her conception of justice be legitimately implementable.
legitimately implementable according to this account in more detail and then
distinguish two strands. The first, the encompassing view, says that claims of
justice encompass claims of legitimacy so that if an action is just, then its
coercive enforcement is legitimate. The ‘adapted’ view says that claims of
reasonable conceptions of justice—which are broader than perfect justice—
establish the boundaries of legitimately implementable actions. I test JAF
against each of these accounts of legitimacy. The first is implausible for RU
theorists, but the second can be adopted to show a conception of justice to
meet the legitimacy requirement.

The natural duty of justice explains political legitimacy in the following way.
Proponents assume, along with Rawls, Kant, and others that we are each under
a natural duty of justice ‘to support and to comply with just institutions that
exist and apply to us’ (Rawls, 1999, p. 99). The idea is that since the purpose
of the state is to secure and maintain just conditions, citizens can fulfil their
duty of justice by adhering to its orders. The state has the right to exercise
coercive control because when it does so it enables each citizen to better fulfil
the duties of justice she is under (Quong, 2011, pp. 130-1). This view is
expressed as a sufficient, but not necessary, condition of legitimacy; it does not
preclude the existence of other plausible sources of the right to coerce people.

According to the encompassing view, justice subsumes legitimacy so that the
state may permissibly coerce citizens into meeting the requirements of justice.
Proponents of this view might agree that legitimate implementability is an
important feature for a conception of justice, but reject the idea that a separate
account of legitimacy is necessary to show the circumstances in which coercion
is permissible. If justice requires me to save a drowning child, then it would be
just for anyone to force me to do so. If justice requires that I give some of my
wealth to the least well-off, then the state can permissibly force me to do so
(the legitimacy right is exclusive to the state in this case if only it can facilitate
justice—in this case by finding the worst-off and redistributing, for example).

---

according to this account. I do not wish to claim that it’s impossible for an RU theorist to
succeed in showing this. However, the fact that the normal justification thesis is not typically
deployed as establishing general conditions of permissible coercion gives me a reason to refrain
from considering it further here.
On this view a conception of justice is legitimately implementable simply in virtue of its being the right conception of justice. If JAF is the right conception of justice, then ipso facto it is legitimately implementable.

RU theorists should reject this account of legitimacy because it’s insufficiently sensitive to the fact of reasonable pluralism and because it yields counter-intuitively narrow constraints to legitimate state actions. Let’s consider how this account would work as a legitimacy test for JAF. Even if we think that JAF is the true conception of justice, there are and always will be people who reasonably disagree (as we saw in the previous chapter). It is unclear that the truth of JAF is sufficient to justify the infringement of their rights to self-govern if we coerce them into complying with its directives. As Valentini says, in circumstances of reasonable disagreement, a state cannot treat citizens justly if it simply imposes one, reasonably contestable, view of justice on them (2013, p. 193).55 Citizens might plausibly come to different conclusions about whether the principles of JAF permit, for instance, a perfectionist sugar tax. Arguments based on values found in the public political culture (as RU justice requires) can be made both for and against the tax and so this disagreement is reasonable. The encompassing view implies that states can only coercively enforce laws that are perfectly just, so if a democratic or some other just procedure yields an unjust result, then it cannot confer legitimacy (Quong, 2011, p. 131). Let’s say the truth is that justice prohibits the sugar tax. Even if every citizen of a given society reasonably disagrees and thinks her duties of justice would be better fulfilled by implementing the tax, there is no possible way for the state to legitimately do so on this view. This result is counter-intuitive. Whether JAF permits the sugar tax, or any such legislation about which there is reasonable disagreement, is ‘a difficult decision and surely a liberal democratic state could legitimately exercise power in a number of different ways in such a case’ (ibid., p. 132). Estlund also rejects the encompassing view because it fails to meet his acceptability requirement. That

55 For Valentini, treating citizens justly means showing them equal respect (ibid.). We can understand this requirement as the grounds of the prima facie wrongfulness of coercion and conclude, with Valentini, that this implies that the state cannot coercively enforce a particular contestable conception of justice onto people in circumstances of reasonable disagreement.
is, accounts of legitimacy that ‘say a law is legitimate simply because it meets the independent standards of justice, will not have a justification that is acceptable to all qualified points of view’ (Estlund, 2008, p. 99). Some qualified views disagree that the sugar tax is a requirement of justice and this undermines the justifiability of coercive actions to force people to comply. That said ‘we cannot suspend judgement and refrain from taking decisions about social distributions until full agreement on matters of justice has been reached, as this would lead to social paralysis’ (Valentini, 2013, p. 193). These objections to the encompassing view are particularly relevant to RU theorists because their conceptions of justice must be sensitive to the fact of reasonable pluralism. Utopian theorists, on the other hand, might idealise away such disagreement and thus consider this a plausible account of legitimacy. Moreover, RU theorists must produce a conception of justice that is sufficiently determinate and so reasonable pluralism cannot result in the state remaining at an impasse with regards to how to act.

Quong’s alternative—what I call the adapted account of the natural duty of justice—is to understand acceptability to reasonable conceptions of justice as a constraint on legitimacy (ibid. p. 132), rather than taking the perfect conception of justice to encompass it. This means that an action ‘may not be [perfectly] just and still be legitimate, provided it is just enough in view of the circumstances and social conditions’ (Rawls, 2005, p. 428). Taking reasonable conceptions of justice to establish constraints in this way allows Quong to adapt the natural duty of justice account so that it is compatible with the circumstances created by reasonable pluralism. He appeals to reasonable, or public, conceptions of justice as the grounds of legitimate coercive actions without defending the claim that those conceptions are true, or perfect, justice (ibid. p. 135). Theorists can show their conception of justice to be within these boundaries, but supplement it with the extra moral reasoning that they think makes it the best/true/most reasonable conception. On the adapted view, the concept of justice does not encompass the concept of legitimacy, but rather establishes constraints that multiple conceptions of justice lie within. In this way the adapted account imposes fewer restrictions on legitimate laws than the encompassing account. It says that a conception of justice is legitimately
implementable if its requirements are within the constraints of the reasonably just. The state can legitimately coerce people to comply with any reasonably just laws in order for them to best fulfil their natural duty of justice.

As we saw in the previous chapter, public reason informs reasonable conceptions of justice. In line with this, Quong adopts Rawls’s *liberal principle of legitimacy* to establish the limits of reasonably just laws. It states that ‘our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason’ (Rawls, 2005, p. 137). This principle ensures that legitimate actions are acceptable to all qualified points of view, as Estlund requires. The appeal to ‘common human reason’ makes legitimacy bounded by public reason; ‘power is exercised legitimately or properly only when its exercise is based on what the guidelines of public reason say are the right kinds of considerations’ (Weithman, 2011, p. 314). Although the content of legitimacy constraints is given by public reason, the natural duty of justice still grounds it. The state’s actions are legitimate when they enable citizens to better fulfil their natural duties of justice. The difference between the adapted and encompassing view is that the former ties legitimacy to the spectrum of possibilities within the constraints of public reason, rather than to the particular requirements of a conception of justice.

To test JAF against the adapted account, we must consider whether its requirements are within the constraints set by public reason in the society to which it will apply. I argued in the previous chapter that RU conceptions of justice must be supportable by public reasons in order to meet the stability requirement. I have stipulated that JAF meets the stability requirement and as such it is also legitimately implementable according to this account. The constraints of legitimate state action are set in the same way as the limits of public reason. So, if a conception of justice is shown to be stable for a particular constituency, then it is also legitimately implementable.

---

54 Although Estlund himself does not commit to a substantive position regarding whether reasonableness is the best measure of a view being ‘qualified’ (Estlund, 2008, p. 63).
One question that we must raise with regards to the adapted natural duty of justice account is whether we can preserve a justificatory role for legitimacy if we adopt it. Recall that legitimate implementability is to contribute to the grounds of RU conceptions of justice. If RU theorists adopt this account then legitimacy will never affect the content of a conception of justice since stability and legitimacy impose the same constraints in this case (both are established by public reason and reasonable justness). The content of justice is already within the constraints of the reasonable in order to meet the stability requirement. However, to say that justice and legitimacy both have a common basis in fundamental political values does not imply that they impose the same demands, or that one reduces to the other (Peter, 2009, p. 59). Both concepts inform one another but the roles are distinct and so the relation is not circular in a problematic way. Justice informs legitimacy by defining the boundaries of legitimate coercive actions in terms of reasonable conceptions of justice. This account of legitimacy informs justice on the other hand, by explaining in what circumstances states are morally permitted to coercively implement a particular conception of justice. So this account of legitimacy contributes to the grounds of a conception of justice by showing it to be legitimately implementable (which is necessary for RU theory because such implementation is always coercive), but it does not inform its content (in any way that differs from the stability requirement). The content of a conception of justice is given by whichever moral reasoning the RU theorist adopts and is compatible with the core commitment of the approach. Legitimate implementability contributes to the justificiation of JAF (and other compatible conceptions). Given an alternative to the natural duty of justice, it may also constrain the content. Here, the appeal to public reasons means that the content remains the same.

The purpose of this discussion was to consider whether the RU approach to justice is compatible with the natural duty of justice account of legitimacy. We can see now that RU theorists should reject the encompassing view, but that an

55 As we shall see, alternative accounts of legitimacy don’t rely on the natural duty of justice and as such constrain the content of substantive principles of justice. The content of justice is not similarly constrained on this account because the shape of public reasons is already given by meeting the stability requirement.
adaptation of the account can be adopted to meet the legitimacy requirement. It is not my aim to suggest that this is the only, or most, plausible account of legitimacy. There are several more that I am yet to consider, each of which presumes a distinction between the justness of an action and the legitimacy of coercively enforcing it. Justice and legitimacy are most famously distinguished in this way by A. John Simmons. Simmons points out that the justice or goodness of a state may give us a moral reason not to undermine it, but that it does not follow from this that the state has an exclusive moral right to direct, be obeyed by, and coerce us (1999, p. 768). The distinction between the two values is said to be as follows. Justifications appeal to an institution’s virtues, goodness, or other beneficial qualities (Quong, 2011, p. 109). Those qualities and benefits vary with conceptions of justice depending on the requirements of particular methodologies. In this chapter I argue that legitimate implementability of a conception is itself a beneficial quality. A theory’s implementation is legitimate, on the other hand, if it meets the requirements of any plausible legitimacy account(s).  

4.2.1.2 Natural Duty of Rescue
This account of legitimacy starts with the assumption that membership in political society is beneficial to citizens and uses that fact as a premise in an argument to show that legitimate coercion can be grounded in the natural duty of rescue. Proponents of this view assume that membership in political society is beneficial compared to the alternative of the state of nature. The natural duty of rescue requires citizens to rescue others from peril when they can do so at little cost to themselves. For example, it’s quite clear that we have a natural duty to save a drowning child if the only cost to ourselves is to be cold and wet for a while. Christopher H. Wellman (1996) has outlined the argument from a duty of rescue to political legitimacy the most clearly.

The essence of Wellman’s claim is that ‘the state’s right to coerce its citizens correlates to the latter’s lack of rights to be free from political coercion, and this absence of rights stems from their Samaritan duties to others’ (1996, p. 56).

Theorists may adopt more than one legitimacy account. Each can partially or wholly explain the permissibility of coercive implementation of a conception of justice.
Our Samaritan duty (to rescue others from peril) is relevant in the political context because being in the state of nature is a type of peril. Wellman describes the nature of this duty thusly: ‘it requires one to assist others only when such assistance is not unreasonably costly and will save the person from peril’ (ibid., p. 235). If we were to secede from political society, then other citizens would face this peril. As such, we lack the right to do so since being a member of political society is of little cost to ourselves. If we lack the right to secede, then the state has the liberty right to coerce us into complying with the demands of political society. The source of legitimacy is a natural duty because it is the existence of this duty of rescue that puts citizens into the position whereby the coercion is permissible. So, the prima facie wrongfulness of coercion is compatible with the implementation of a conception of justice when citizens’ membership in the coercive relation is required as a result of their duty of rescue. The existence of each citizen’s moral duty to others allows her state to coerce her in order to secure political stability for her fellow citizens (ibid., p. 219). According to this account, states must exist because the moral rights we’re concerned about in the state of nature will only be respected ‘if police effectively protect individuals and recognised judges impartially adjudicate conflicts according to established rules’ (ibid., p. 217). This is a plausible account of the circumstances in which coercive state action is justified.

Let’s see how we might go about testing JAF against this account. On this view, a conception of justice is legitimately implementable if citizens must heed its requirements in order to fulfil their natural duty of rescue. Recall that Wellman describes the nature of this duty as requiring one to assist others only when such assistance is not unreasonably costly and will save the person from peril’ (ibid., p. 235). As such, we must consider whether the requirements of JAF are 1) necessary to save other citizens from peril, and 2) not unreasonably costly. The first principle—that stipulates a set of equal basic liberties for all—meets these requirements. Infringement of basic rights is exactly the way that Wellman describes the peril of the state of nature and so enforcing the first principle is necessary in this sense. We can also presume that doing so will not be unreasonably costly since the equal distribution means that the benefits and
bureaucrats of such a scheme are shared by all. Moreover, the special conception of justice specifies that the priority of this principle only applies in context where there are sufficient resources to secure it. So, the first principle of JAF is legitimately implementable. It’s unclear whether the same is true of the second. The second principle requires fair equality of opportunity (as well as the difference principle). Securing fair equality of opportunity to offices and positions is not necessary to avoid peril. The fact that I’m ineligible to apply for a particular job because I am a woman is a morally concerning inequality, but it does not necessarily leave me in peril. In addition, the requirements of fair equality of opportunity may be somewhat substantial (requiring the state to provide healthcare, education, and training for instance) and could therefore impose significant costs. It’s unclear that contributing to the provision of fair equality of opportunity is something that citizens lack the right to secede from their involvement in. In this case a citizen is not failing to fulfil her duty of rescue if she doesn’t pay. If the second principle is either unnecessary to avoid peril or unreasonably costly, then the second principle of JAF is not legitimately implementable according to this account. The RU theorist must then accept that only a portion of JAF can be justified by its legitimate implementability and reject or revise the other portion (unless she establishes the legitimate implementability of the other portion via an alternative account).

As will be clear from the test of JAF above, this account of legitimacy implies an extremely narrow scope for legitimate state actions. The existence of a duty to rescue is plausible, but the only coercive action it legitimates is the bare minimum needed for citizens to avoid the perils of state of nature. All the familiar existing RU conceptions of justice require a broader range of coercive state actions than those this account permits. This isn’t to say that this account of legitimacy is fundamentally incompatible with the RU approach. As we have seen, a conception that consisted of the first principle of JAF would be legitimately implementable. However, since the range of actions available to states is so narrow on this account, many theorists may find it to be implausible or in need of supplementing. It’s a plausible a priori account of legitimacy, but it fails to offer much help to the RU theorist in terms of demonstrating her conception of justice to be legitimately implementable.
I now turn to consider an alternative to natural duties as the grounds of legitimacy: the idea that it is the result of an action in which the state acquires the right to rule.

4.2.2 Acquired Legitimacy

Acquired rights and duties result from specific interactions. Some accounts of legitimacy explain the permissibility of coercive state interference by demonstrating the existence of its acquired moral right to perform such actions. The prima facie wrongfulness of coercively implementing a conception of justice is not problematic if the state has such a right. There are several suggestions of the particular interaction(s) that results in this right. I discuss the most prevalent suggestions of the way that the state secures it: as a recipient of consent to be ruled, as the recipient of citizens’ obedience out of a duty of fairness, and as the imposer of the result of democratic procedure. Each of these interactions requires particular substantive or procedural features of a conception of justice to be true in order for that conception to be legitimately implementable. I conclude that democratic authority is a plausible account of legitimacy for RU theorists, but that consent theory and fair play are not.

4.2.2.1 Consent Theory

The claim that consent can grant the right to coerce is widely accepted. Its intuitive appeal derives from the common-sense idea that I can transfer or waive my rights (including the right to self-govern or whichever right grounds the prima facie wrongfulness of coercion) to someone else via an act of consent. If we have the Hohfeldian power to transfer our right to self-govern, then our consent to the state (or whomever) is not wrongful because coercion does not infringe our right to self-govern. Imagine, for example, that Bill desperately wants to eat healthily but lacks the knowledge or the willpower to do so (or to learn what would constitute healthy eating). Bill’s friend, Ted, is a nutritionist and so can help Bill in this regard. Bill asks Ted to do so and gives Ted full responsibility for his diet, and asks Ted to punish him if he deviates. In this example Ted can legitimately tell Bill what to do and use coercive force to get him to comply. Because Bill consented to giving up his right to self-govern, there’s no wrong when Ted coerces him in this way. Citizens can
similarly consent to the state. So, ‘political power is morally legitimate…only where the subjects have freely consented to the exercise of such power and only where that power continues to be exercised within the terms of the consent given’ (Simmons, 1999, p. 745).

Clearly, not every putative act of consent can authorise a rights transfer of this kind. If Bill consents to Ted’s control because of misleading information from others (perhaps his diet is actually fine, but he’s been misled into believing it’s unhealthy in order to sell him Herbalife products), then his consent oughtn’t to result in a rights transfer. Similarly, if he faces dire consequences by not consenting (perhaps Ted is a control freak and said that he will shoot Bill if he doesn’t let him take charge), then we shouldn’t take Bill’s consent to have legitimately transferred his right. In order to avoid these difficulties we accept that only an appropriate act of consent binds the consenter in the relevant sense. Although it’s clearly sensible to reject examples like the ones just given, theorists disagree over the exact specification of the conditions of appropriate consent. It’s not germane to our task to discuss these here so I take the following three to reflect our most prominent intuitions: the consenter must act autonomously, have access to suitable alternatives, and be aware of the consequences of her consent in order for it to ‘count’ as an appropriate act.

Appropriate consent is, of course, a plausible justification of coercion, but RU theorists cannot adopt this account as a necessary condition of political legitimacy because its satisfaction could never be achieved in any version of the world as we know it. Imagine, once again, that we are testing JAF, an RU conception of justice, to see whether it is legitimately implementable. According to this account, the conception is only legitimately implementable if the constituency to whom the conception applies have consented to being governed by it. Although they consent to the actual state (as opposed to the conception), we can presume that they will judge states according to the conception that guides their actions. So, we must consider whether everyone in the relevant constituency could simultaneously consent to the particular requirements of JAF. The relevant constituency for RU theorists, as we know, is reasonable people. Now, although the requirements of JAF are compatible
with the values of a reasonable person (so much is necessary to meet the stability requirement), their being so is insufficient to demonstrate that she would consent to it. The requirement that citizens have suitable available alternatives to consenting to the state, for instance, requires unlikely empirical circumstances and, according to Buchanan, makes consent theory ‘remarkably ill-suited to the political world and so extraordinarily demanding as to be utopian in the worst sense’ (Buchanan, 2002, p. 699). Even though the content of JAF is such that reasonable people have sufficient reasons to agree with it, the conditions required for their consent to be binding cannot be secured in an RU idealised society. Recall that RU theory must be implementable into a version of the world as we know it. As such universal consent is chimerical and justification of coercion cannot feasibly require de facto unanimous consent (Valentini, 2012, p. 596). Since consent requires all citizens to simultaneously agree about which state and laws they may be coerced into complying with and for that agreement to be made even when suitable alternatives are available, ‘nothing like what we now call a government satisfies or is ever likely to satisfy the conditions required for all or even most of its citizens to consent to its exercise of political power according to any conception of consent that would justify the exercise of political power’ (ibid.). As such, this account of consent is incompatible with the other methodological requirements of the RU approach.

Given this difficulty, some proponents of consent theory have suggested that there are alternative actions that can serve as a proxy for consent and ground legitimacy in the same way (for example: voting, using state services, or residing within the state’s territory). However, these alternatives fail to meet the conditions that make the act of consent appropriate. Although I haven’t conclusively defended those conditions and thus ruled out this possibility, these alternatives are not sufficiently plausible to warrant further discussion. If a conception of justice can be shown to meet the requirements of this account, then it would certainly contribute to the justification of that conception. However, the theorist’s commitment to producing a conception that could
govern a version of the world as we know it means that this account is unsuitable for RU theories.  

4.2.2.2 Fair Play Theory

Another account of legitimacy that grounds the state’s right to coerce its citizens as a correlate of citizen’s acquired duty to obey is fair play theory. Fair play theorists cite a duty of fairness that we acquire in certain circumstances. The classical statement of the view, from H. L. A. Hart, is the following:

When a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions when required have a right to a similar submission from those who have benefited by their submission (1955, p. 185)

This duty requires people who benefit from the cooperative efforts of others to themselves cooperate in the joint enterprise that produced the benefits. Benefiting from others’ efforts without contributing yourself, or ‘free-riding’, is commonly thought to be morally wrong. For example, if a group of housemates take turns to cook dinner and clean-up, then you would be doing something wrong if you ate the dinners and enjoyed the clean kitchen, but refused to help in either activity. Fair play draws on that intuition and claims that we have a natural duty of fairness not to free-ride. In the political context, if the existence of the state benefits you, and citizenship and obedience to the state is the right kind of cooperative enterprise, then failure to obey the law is unfair in this way. Applying these conditions to the political context is plausible since ‘there can be little doubt that the citizens of many modern states enjoy nonexcludable public goods that depend upon the cooperative efforts of their fellows’ (Klosko, 1987, p. 356). Obedience to the state is necessary in order to avoid free-riding and fulfil your duty of fair play. So, this account says that the state’s right to command citizens correlates with their duty to obey which they

---

37 Even if it only imposed very minimal requirements that no rational person could deny the value of, the presence of any ‘suitable alternatives’ will invoke disagreement over which preference so that universal consent is not possible.
owe as a result of their duty of fair play. I consider the leading proponent George Klosko’s account (2004) of fair play theory in further detail.

The main obstacle to fair-play theory being a plausible explanation of political legitimacy is the difficulty in explaining why the duty of fairness is enforceable (and only enforceable by the state and not by other citizens). It’s unclear that the duty not to free-ride is sufficiently strong to be enforceable. Without an explanation of the right to force citizens to comply with their duty to obey the law, accounts of political obligation may successfully explain the source of the right to issue commands (the state’s authority), but do not show why the state has the right to force citizens to comply with those commands (its legitimacy). Since I have claimed that the coercive nature of legitimacy is what makes it relevant to RU theory, the enforceable nature of the duty to obey the law is vital for RU theorists to fulfil the legitimacy requirement. Klosko says that the duties are enforceable because non-cooperation creates collective harms that are objectionable on the basis of the harm principle (2004, pp. 45-8). Drawing on the famous principle that Mill advocates in *On Liberty* (1869), Klosko argues that there are at least some collective patterns of behaviour that are necessary to the well-being of all members of society. Examples of requirements entailed by such patterns of collective behaviour include citizens paying taxes, contributing to anti-pollution initiatives, and giving evidence in a court of law. States can legitimately force citizens to cooperate into doing these things because when they fail to cooperate they are undermining the well-being of their fellow citizens and thus violating the harm principle. Violation of the harm principle is, for Klosko, the clearest case in which infringement of citizens’ liberty (their right to self-govern) via coercion is justified. As such, citizens’ duty to cooperate is enforceable when their non-cooperation would violate this principle.

Adherence to the harm principle makes the source of legitimacy (the enforceability of the duty to cooperate and obey the law) a duty of justice, and not fairness. It’s permissible for the state to coerce citizens’ into cooperating with the law because when they fail to do so they violate the harm principle (a principle of justice), and not because they have a duty of fairness to obey. In
this way, the relevant test for RU theorists is to consider whether citizens must adhere to the requirements of their conception of justice to avoid violating the harm principle. There’s no need for them to consider the duty of fairness unless they have an independent reason to be concerned with the moral source of the duty to obey. Of course, whether or not the requirements of JAF are necessary in this way depends on the particular interpretation of harm, something that is ‘fraught with complexities’ (ibid., p. 46). However, given the existence of the kind of collective harms discussed above, it is plausible that the requirements of JAF could be legitimately implementable. Failing to pay the taxes required for fair equality of opportunity, or oppressing someone on the basis of her religion (a violation of her basic liberty), for instance, can both plausibly be understood as inflictions of harm, so long as the concept is sufficiently broad. A narrower interpretation of harm would, of course, restrict the range of legitimately implementable state actions. I do not defend any particular interpretation of the harm principle here, but simply point out that it could be a plausible source of political legitimacy since it’s compatible with the other requirements of the RU approach. Fair play theory, on the other hand, is insufficient to explain the permissibility of coercively implementing a conception of justice.

4.2.2.3 Democratic Legitimacy

According to this account, the state acquires the right to coerce people into obeying particular laws and policies if those laws and policies are the result of a democratic procedure. ‘Whatever the results of discussion, deliberation, and decision-making are, they are legitimate. The results are made legitimate by being the results of the procedure.’ (Christiano, 1996, p. 35) There is a vast literature surrounding the value of democracy and particular democratic features necessary for the results to be legitimate. I understand a general notion of democracy to refer to ‘a set of real-world collective decision-making processes in which those who belong to a particular group (society in the case at hand) have a right to an equal say in establishing the rules that apply to them’ (Valentini, 2013, p. 179). RU theorists may specify and defend particular democratic features they find plausible (deciding, for example, whether proportional representation or a ranked voting method is preferable). Given
disagreements about such features, I leave open the question of which does the 
best job of establishing legitimacy. I briefly discuss the essentials of democratic 
proceduralism in order to explain how democratic procedures can be a source 
of legitimacy, and make some brief remarks about the way in which the general 
democratic account could be compatible with the RU approach.

Whilst all views of democratic legitimacy take proceduralism to be important, 
theorists disagree over whether additional conditions must be met in order for 
the results of the democratic procedure to be legitimate. The Pure 
Proceduralist view says that democratic decisions are the result of equal 
considerations of individual interests only (Peter, 2007, p. 332) whereas some 
accounts add an epistemic component which stipulates democratic procedures 
to confer legitimacy because they yield results that are correct by some 
independent standard. Rather than distinguish and discuss all such variations, I 
briefly outline Estlund’s view, which is a combination of the two (‘it combines 
the normative appeal of ‘correctness’ with that of fair procedures’ (ibid., p. 
339)).

According to Estlund’s account a necessary condition for the state to 
coercively enforce a law is that it is the result of the right procedure (Estlund, 
2008, p. 41). Of competing procedural sources, ‘democracy will be the best 
epistemic strategy from among those that are generally acceptable’ (ibid., p. 
42). So, on his view, we should adopt a procedure that is both i) likely to make 
the correct decisions (in terms of justice) and is ii) generally acceptable to all 
qualified points of view. Democracy has both of these features; it tends to 
produce decisions that are good from the perspective of a public, or true, 
conception of justice (ibid., p. 112) and it is generally acceptable to all qualified 
points of view because it treats people equally. The epistemic element of 
Estlund’s account is appealing because it helps to alleviate concerns that 
democratic legitimacy can prevent the legitimate implementation of laws or 
policies that we think are otherwise perfectly just. For instance, JAF requires 
fair equality of opportunity, but laws protecting it are only legitimate when they
receive support from a democratic procedure.\textsuperscript{58} If citizens happen to vote for an alternative that only secures formal equality of opportunity, say, then fair equality of opportunity, that we otherwise took to be a requirement of justice, is illegitimate in this context. If Estlund is right about the epistemic qualities of democratic procedures, then the results of this example are less troubling since we can be sure that whichever alternative they selected (in this case formal equality of opportunity) has some independent moral worth and they won’t vote for something obviously unjust (like offices and positions being open to white people only).

There are some requirements of a conception of justice whose legitimacy cannot be established by democratic proceduralism. Inevitably, the nature of the procedure itself cannot be democratically determined. This is true of other constitutional essentials, too (there needs to be some idea of how institutions are to be arranged before the imposition of a democratic procedure is even possible). For this reason, theorists must supplement this account of legitimacy to explain why it is permissible for the state to coercively enforce the aspects of their conception that cannot possibly be the result of a democratic procedure. The adapted version of the natural duty of justice is a good example of an alternative account that could supplement democratic legitimacy in this way. It specifies reasonable justice as the constraint on the content of legitimately enforceable laws and says that states have the right to coerce citizens because being under coercive control is a better way for citizens to fulfil their duty of justice than acting individually. Treating justice as constraining (rather than encompassing) legitimacy leaves the precise content of many laws indeterminate and merely stipulates that they must not conflict with reasonable conceptions of justice. Theorists may then appeal to democratic procedures as the source of legitimacy to coercively enforce particular laws and policies within the constraints of the reasonably just. So, the state may permissibly coerce citizens \textit{at all} because of the natural duty of justice they are under, and theorists may determine the particular content of permissible actions by

---

\textsuperscript{58} This is distinct from whether they are legitimately implementable, the conditions for which I discuss shortly.
appealing to just procedures, which are likely to be democratic. This combination of legitimacy accounts is also attractive because it limits the power of the demos so they can’t legitimise actions that we take to be clearly unjust according to other grounds (for example, those that violate the first principle of JAF to equal basic liberties). The legitimacy of constitutional essentials could be established by other accounts, I merely use the natural duty of justice as an illustrative example to explain one such possibility here.

According to this account, a conception of justice is legitimately implementable if its requirements could receive support as the result of a democratic procedure. It need only be possible since RU conceptions are designed to govern idealised societies. So the question of whether they are legitimate in a particular time and place in the actual world is not relevant. Since legitimacy has a justificatory role, this means that the potential results of democratic procedures form a part of the grounds of RU conceptions of justice. We can, once again, run the legitimacy test against JAF. In order to show it to be legitimately implementable, we must show that JAF could be the result of a democratic procedure in the society it is intended to govern. Since the conception is intended to govern a constituency of reasonable people, the possibility of JAF being the result of a democratic procedure is established by its stability. If citizens have sufficient reasons to be motivated to comply with the requirements of JAF, then a majority could plausibly support its adoption via a democratic procedure. Since democracy doesn’t require the express endorsement of each individual citizen, this requirement is plausible where consent theory was not.

A notable implication of this account of legitimacy is that it imposes a substantive demand on the content of RU theories by requiring them to adopt a democratic constitution. Clearly, such a constitution is compatible with the

59 Just procedures are likely to be democratic because of substantive commitments to freedom and equality reflected in many conceptions of democracy. Indeed many ‘contemporary liberal theorists agree that, to be consistent with equal respect, a society must be democratically organized, giving its citizens equal rights to participate in political decision making’ (Valentini, 2013, p. 177). The details of such an argument are beside the point here.

60 And they meet the requirements of the supplementary account when it comes to constitutional essentials.
requirements of JAF (Rawls presumes such a constitution in the moderately idealised society to which JAF applies (2005, p. 3)). The results of the democratic procedure shaped by JAF then inform the particular laws and policies that the state can coercively enforce. Moreover, given the values that both epistemic and procedural accounts of democratic legitimacy are based upon (justice, fairness, freedom, and equality), this will be a plausible commitment for any RU conception of justice so-long as those values are found in the PPC of the constituency the theorist intends her principles to govern. That is, the substantive commitment to some form of a democracy imposes a real constraint onto the content of RU conceptions of justice, but is likely to be one that most can accommodate because of the values it is based upon. If a conception of justice is not compatible with democracy in this way, and the theorist adopts this account of legitimacy, then the conception must be revised so that a democratic constituency is possible and the conception can be shown to be legitimately implementable.

An RU theorist may adopt any account of legitimacy that I have shown to be compatible with her approach here. She may even develop or adopt one that I have not considered. The purpose of this section has been to consider the implications of adopting particular accounts. Whichever account the RU theorist chooses influences her methodology, and sometimes the substance, of her theory in different ways. Having settled on a plausible account, the RU theorist must then show that the content of her conception of justice is compatible with its requirements as I have begun to do with the requirements of JAF as an example here. Then, she may claim that the conception is legitimately implementable and this status forms part of the grounds for adopting it over alternatives.

4.3 People’s Views in Accounts of Legitimacy

The purpose of this chapter is to establish the claim that RU conceptions of justice should be legitimately implementable and to investigate the impact of this requirement on the plausibility of existing accounts of legitimacy. Some such accounts, notably the adapted natural duty of justice and democratic
proceduralism, are clearly compatible with the requirements of the RU approach, including legitimate implementability. At the heart of both is a claim that justifiability to citizens is in some sense important for legitimacy. The role of a public conception of justice demonstrates this in the former, and the role for the people’s voice in a democracy demonstrates it in the latter. As such, and in line with the structure of previous chapters, I now turn to consider whether the legitimacy requirement gives RU theorists another reason to suppose that what the people think matters for the content of a theory of justice.

Recall that when I discuss people’s views I have in mind their views about justice, morality more widely construed, or in this case perhaps legitimacy. In previous discussions of people’s views, I considered their expression to be via: a coarse representation via surveys or voting; the social meanings of goods or practices; the relations they exhibit with one another; or one of the various interpretations of the public political culture (PPC). I concluded at the end of the previous chapter that if theorists allow actual people’s views to inform the PPC, and they accept the consensus or revised view of stability, then actual people’s views have a role in the justification of substantive conceptions of justice. Giving people’s views this role allows theorists to simultaneously meet the weak contextualism requirement and the stability requirement. As a result of this possibility, sensitivity to people’s views is an attractive proposal for RU theorists and a unifying theme of the methodological requirements I’ve defended. The reasoning in favour of it is strengthened by some of the accounts of legitimacy I consider in this chapter. I return first to the adapted natural duty of justice account.

If theorists adopt this account, then people’s views inform legitimacy in the same way that they inform stability in the previous chapter. The account appeals to reasonable conceptions of justice to establish the boundaries of legitimately implementable laws and policies. We saw in the previous chapter 61 Valentini has interpreted a commitment to equal respect to require that legitimacy demands ‘institutional mechanisms which ‘force’ a certain distribution of rights on people be as consistent as possible with people’s own views’ (2013, p. 597).
that conformity to reasonable conceptions of justice requires theorists to appeal to the PPC as the source of public values that inform such conceptions and that actual people’s views necessarily inform these values. Since reasonable conceptions of justice also constrain principles’ legitimacy, we can say that it too is informed by people’s views in the same way that stable conceptions of justice are. ‘Only a political conception that all citizens might be reasonably expected to endorse’ (Rawls, 2005, p. 137) can justify the coercive use of state power. This account could prescribe a political conception of justice in which people’s views have an even stronger role. It might require the precise content of the state’s coercive actions to be a function of the views to those to whom it will apply. In this case not only do people’s views inform the justification of legitimate actions, but they inform the content of them via a democratic procedure.

If the results of democratic procedures have a constitutive role to the content of justice, then people’s views have a more direct role than I have so far considered. If RU theories of justice include a democratic constitution—because such a constitution is necessary to establish legitimacy, or because it is a substantive requirement of justice—then RU theorists have another reason to acknowledge a role for people’s views in a theory of justice since ‘citizen-responsiveness is at the heart of democratic politics’ (Valentini, 2012, p. 606). Regardless of the particular democratic procedure theorists adopt and its ability to accurately map what people want, it is indisputable that the results of a democratic procedure are in some way influenced by the views of those participating.

Although, as I argue in chapter two, RU principles of justice must be determinate to some extent, they are not required to specify institutional arrangements beyond those necessary to establish whether a conception is ultimately implementable. So, whilst I have said that RU theorist’s idealised societies must be organised by a state and resemble a version (and area) of the world as we know it, I have not specified any particular institutional arrangements to be necessary to establish the content of the PPC and meet each requirement. In this chapter, I considered the proposal that a democratic
procedure is, in fact, a particular institutional arrangement required to establish legitimacy. It may be the case that democratic procedures can also enable a conception to meet the stability and contextualism requirements. The stability of a conception of justice is informed by its compatibility with the values reflected in the PPC. A social scientist may uncover the values implicit in the PPC without relying on the results of democratic procedures. However, once a conception of justice has been implemented, the theorist must ensure that her conception remains sensitive to the values implicit in the PPC in order to continue to meet the stability requirement. Democratic procedures are a good way to do this. They can track changes in attitudes and allow policies to respond. The conception of justice that governs a society can be malleable according to the results of the procedure. In this way, RU theorists might adopt a democratic system for stability reasons and independently of whether such a system also grounds legitimacy. As we have seen, if a theory meets the stability requirement in this way, then it will meet the contextualism requirement simultaneously.

The other compatible accounts of legitimacy that I discussed may not give theorists additional reasons to consider people’s views. This fact does not detract from the role I established for such views in the previous two chapters. It may be the case that the justificatory contribution of legitimate implementability is not informed by people’s views, but that contextual and stability ones are. We might then return to consider whether any alternative types of contextual facts from the second chapter can meet each requirement in the same way that people’s views about justice can. That is, whether it’s plausible that another type of contextual fact can meet the weak contextualism, stability, and legitimacy requirements simultaneously (or whether there are distinct but compatible means to meeting each of them separately). The types of contextual facts that I considered apart from people’s views include facts about the economic situation, natural resources, existing institutions, infrastructure, and history. Whilst the conception could meet the weak contextualism requirement via sensitivity to any of these things, they could not simultaneously show the conception of justice to be stable for the right reasons. This is not a problem in itself since simultaneity is not a requirement.
However, stability requires theorists to show that reasonable people could be motivated to comply with the requirements of a procedure and it’s unclear that theorists could do this without appealing to the views of people via the PPC. Since this is the only way for theorists to meet this requirement, and these facts will always be contextual, there’s no need to require sensitivity to another type of contextual fact. As such, sensitivity to people’s views is a necessary condition of an RU conception of justice.

Conclusion

In this chapter I sought to establish a third methodological claim concerning the grounds of RU conceptions of justice. I argued that the legitimate implementability of a conception of justice contributes to its justification. Legitimate implementability is important for RU theories of justice because of the permanence and nature of political power. These facts reveal that coercion is always necessary to implement a conception of justice in any version of the world as we know it. This fact is permanent and the prima facie wrongfulness of coercion makes it problematic for a conception of justice. When a conception of justice can be shown to be legitimately implementable (via an account of legitimacy), then this problem is solved; the coercion the state exercises to implement it is not wrongful.

The idea that legitimacy contributes to the justification of a conception of justice in this way raises interesting questions about the nature of the relation between the concepts of justice and legitimacy. Two prevalent views are incompatible with my claim. First, the idea that justice and legitimacy are completely separate (so that the legitimacy of a conception could not contribute to its justification) was rejected for the reasons given above. Second, the idea that justice encompasses legitimacy (so that an account of the latter would be redundant when developing a conception of justice) is incompatible with this claim. Fortunately, RU theorists have reasons to reject the encompassing view on other grounds. Discussing this prompted the question of whether other prominent accounts of legitimacy would be compatible with its justificatory role and the other methodological commitments of the RU
approach. To answer this question I considered several prevalent accounts of legitimacy, tested a familiar conception of RU justice (JAF) against them, and concluded that all but one—consent theory—were compatible in both of these senses.

Confident that the RU theorist could fulfil my legitimacy requirement without having to reject all existing accounts, I considered finally whether this methodological commitment draws support for the suggestion that people’s views are relevant for RU theories of justice. As we saw in the final section, this depends on the account of legitimacy the theorist chooses, but some prevalent views establish a role and give us a further reason to suppose that it matters what the people think.
5. Conclusion

5.1 Meeting the Aims of the Thesis

The focus of the thesis has been the Realistic Utopian approach to theorising about justice. The aim of political theorists who pursue the RU approach is to develop principles of justice that guide the citizens of a hypothetical, but practically possible, world. My central contribution has been to establish the methodological requirements of such an aim. In particular, I aimed to develop reasons in favour of requirements eschewed by some existing proponents. These reasons are internal to the approach; they draw on the aim of RU theory. As such adopting them provides us with a richer and more fully-worked-out definition of what it means to be an RU theorist. I refrain from proposing particular principles—my aim has not been to establish a full methodology—but my claims are required in the sense that they narrow the range of justificatory and methodological principles compatible with the approach. My requirements merely specify some necessary conditions rather than spelling out a complete list of necessary and sufficient requirements.

Although theorists have different reasons for adopting the RU approach over alternatives, its intrinsic value makes it worth considering the proper methodology for. RU theorists are looking for guidance on how, within the limits of practical possibility, we can be as just as we can. Regardless of its intrinsic value, the prevalence of the RU approach in existing literature makes the contribution of this thesis noteworthy. Even if we needn’t give priority to
RU theories, the thesis is valuable insofar as it speaks to the existing approaches and gives us a means to assess and compare them. In this sense its appeal is broad, and of interest even to those who are less persuaded of the value of developing a set of substantive RU principles.

Although there is some consensus over the core commitment to the aim of developing principles for the constituents of a realistic utopia and the bare essentials of the methodology required to produce it, any more stringent methodological requirements are the focus of significant disagreement or indeterminateness in the literature. I claim that the shared core commitment entails each of the more stringent methodological requirements and as such it is important for RU theorists to recognise these requirements.

Having established how the thesis meets the aims set out at the start of the thesis, I now summarise the contribution I make within each chapter. Following this, I consider the concluding remark of each chapter—that people’s views are important for justice—and clarify the impact of this requirement.

5.2 Summary of my Contribution

In the first chapter I start by setting out what I called the core commitment of the RU approach. It defines RU theorists as those who are committed to producing principles of justice that govern the constituents of a realistic utopia. This aim establishes the limits of the approach and distinguishes it from alternatives on the ideal/non-ideal continuum. Realistic Utopianism is distinguished from Utopianism by its commitment to producing principles that are ‘ultimately implementable’ in the world as we know it. Realistic Utopianism is distinguished from Non-Idealism because proponents of the former reject the claim that facts about the world here and now necessarily feature as constraints on the content of substantive principles of justice. Having made these distinctions, I then discuss the way in which existing RU proponents use idealisations but limit them by acknowledging and responding to feasibility constraints. Responding appropriately to feasibility constraints requires RU
conceptions of justice to be sensitive to some empirical facts, and existing theorists broadly agree that facts reflecting the true laws of science, the circumstances of justice, and the reasonable likelihood of success should feature in their account of feasibility. Which facts should contribute to a more substantial account of feasibility is a source of disagreement. At the end of the first chapter I consider some distinct types of facts that theorists might plausibly include in a conception of feasibility. I do this in order to initiate investigation into why theorists might make their principles sensitive to these things rather than to argue in favour of a thicker account of feasibility. These discussions pave the way for the three chapters that followed.

In the second chapter I start by re-considering the idea that contextual facts should inform feasibility assessments. Rejecting this suggestion, I instead conclude that the core RU commitment requires conceptions to be determinate. To say that a conception of justice is determinate is to say that its principles yield definite verdicts about the justness of particular institutional arrangements or citizen’s/state’s behaviour. Determinacy is necessary in order for theorists to make a claim about whether a principle is ultimately implementable. When the grounds of some principles of justice are partially given by contextual facts, then the guidance of those principles is sufficiently determinate. This is because theorists consider these facts as a part of the details of their idealised society in order to test the implementability of their conception of justice. These grounds for the contextualism requirement are acceptable to existing RU theorists because they match those used to defend their conception’s sensitivity to feasibility constraints.

The arguments in favour of contextualism do not depend on a particular account of contextual facts. Having concluded that some RU fundamental principles should be sensitive to at least some contextual facts, I considered whether actual people’s views about justice are a good candidate. At the end of chapter two I tentatively concluded that they are, but that plausible alternatives are also available. The argument supporting people’s views as the appropriate contextual fact is based on the idea that giving those views serious consideration is necessary in order for theorists to treat citizens as equals. This
idea initiated a theme for discussion at the end of each forthcoming chapter. In each, I asked whether the conclusions of that chapter would draw support for the suggestion regarding people’s views and thus bolster RU theorists’ reasons to accept it.

In the third chapter I defend the stability claim. This claim requires RU theorists to develop a conception of justice that can order society and be stable for the right reasons. The reasoning in favour of this claim draws on the grounds supporting the feasibility commitment. It says that it’s implausible to think that the feasibility of implementing a set of institutions is normatively significant but that whether a conception can support their maintenance is not. In order to show a conception of justice to be stable for the right reasons, the RU theorist must appeal to an account of human motivation. She must establish the characteristics a conception of justice must have in order for her to reasonably claim that it will stably order society. After considering the most prevalent suggestions of such characteristics, I conclude that both the consensus and what I call the revised views of public justification are acceptable and compatible with the RU approach. However, although both are acceptable, theorists should prefer the latter because it permits a deeper and more satisfactory kind of justification.

Having defended the stability requirement, I returned to the suggestion that principles’ contextualist grounds can make RU conceptions of justice sensitive to people’s views. The use of the public political culture in both the consensus and revised views of public justification makes RU conceptions of justice sensitive to actual people’s views. Their input is essential to the PPC, given that its shared nature is what makes it normatively significant. This means that if RU theorists adopt a justificatory principle that demands public justification according to the consensus or revised view, then their methodology may meet both requirements simultaneously. Since people’s views are an essential input, this view gives theorists a conclusive reason to acknowledge at least some sensitivity to people’s views.

In the final chapter I argue that RU conceptions of justice must be legitimately implementable. This requires theorists to defend and adopt a particular
account of legitimacy and then show their conception to be compatible with its requirements. When a conception of justice is shown to be capable of meeting the requirements of a legitimacy account, this contributes to its justification. Accounts of legitimacy must have this role in an RU theory of justice because otherwise citizens are wrongfully coerced when a conception is implemented. Wrongful coercion is incompatible with a conception being justly implemented and so an account of legitimacy is necessary. Although I don’t specify which account of legitimacy I take to be the most plausible I do test an existing RU conception of justice against some of the most prevalent accounts in order to consider whether they are compatible with the RU approach as I had characterised it thus far. I conclude that most prevalent accounts are compatible with the RU approach, but that there are some that are unable to fulfil the justificatory role I specify for legitimacy.

At the end of the chapter I returned, once again, to the suggestion that RU conceptions of justice should in some way be sensitive to people’s views. Although the methodological requirements of the legitimacy claim are not conclusive (since it is up to the theorist to argue in favour of what she takes to be the most plausible account), two of the most prevalent draw yet more support for the suggestion. A qualified version of the natural duty of justice and democratic legitimacy both require some input from people’s views.

So, the two main conclusions of the thesis are as follows. First, the methodological requirements of the RU approach to justice are more stringent than proponents of the approach typically acknowledge. Such a failure to acknowledge the stringency of these methodological requirements is sometimes due to a misunderstanding of the aims or fundamental commitments of the RU approach. For this reason the analysis of the core commitment of the RU approach in the first chapter was essential. In other cases, RU theorists may have overlooked the requirements I defend, or failed to recognise their essential nature. The second main conclusion is that the methodological requirements I endorse can be unified by the idea that RU theorists should acknowledge a role for people’s views about justice. These reasons are not conclusive from the weak contextualism and legitimacy
requirements because there are other ways to meet them, but they are a necessary input to meeting the stability requirement. As such, taking people’s views about justice seriously is the most effective way to simultaneously meet all three requirements. Accepting this conclusion may make more of a difference to the way in which theorists fulfil the requirements of the approach than the other methodological requirements in themselves.

5.3 It Matters What the People Think

I want to conclude by clarifying the sense in which sensitivity to people’s views requires existing RU theorists to adapt their approach to developing a theory of justice, and to consider the additional questions this idea raises. The intuition that the views of those to whom principles of justice are to apply are important is widely held. So much can be gleaned by the wide support for concepts such as autonomy, democracy, and public justification. Indeed, as Miller says, most theorists would ‘claim in one way or another to incorporate and systematize existing beliefs about justice in their theoretical constructions’ (1992, p. 556). However, the extent to which the beliefs of actual people in the world today should inform the content of principles of justice for a realistic utopia has not been fully appreciated by existing RU theorists. Indeed, these theorists would usually reject the claim that ‘such beliefs are in some sense constitutive of a proper understanding of what justice demands’ (Swift, 1999, p. 350).

The traditional worry is that whether or not people think a particular substantive moral concept (i.e. desert or merit etc.) is a just distributive criterion in certain circumstances is irrelevant to the question of whether it is, in fact, a just distributive criterion. However, this concern only arises as a result of existing RU theorists’ failure to understand the role I’ve suggested people’s views have in RU methodology. I have suggested three roles that people’s views might play in an RU theory of justice and, even if we accept all

62 Desert is, according to Swift is the ‘most glaring case of a discrepancy between lay beliefs on the one hand, and those endorsed by political philosophers on the other’ (ibid. p. 357ff).
three, this does not give people's views the kind of strong constitutive role that motivates the traditional worry. The roles I have outlined avoid this concern because ‘the extent of the ambiguity, confusion and inconsistency in those beliefs is sufficient to leave a great deal of work to be done, work that may well lead to conclusions differing markedly in content not only from empirical public opinion understood as those beliefs that survey research shows to be regularly and consensually endorsed by the vast majority of ordinary people but also from the more reflective beliefs that would result from the research method’ (ibid., p. 358). As such, there is still a strong and important role within the RU approach for political theory methods more traditionally understood. Given the middle-ground people’s views must thus occupy, established by the roles for people’s views as I have identified them, we might wonder what RU theorists must do to get the relevant empirical information to incorporate them into the methodology.

It is instructive to first re-consider which views in particular we have in mind. When introducing the idea at various points in the thesis I said that theorists should be interested in what people think about justice, morality, or legitimacy. It was out of respect for citizens’ capacity for a sense of justice that the suggestion of sensitivity to people’s views was first made in chapter two. As such, it would be strange for theorists to consider people’s views about their own plans of life, or conceptions of the good. We saw in the stability chapter that people’s views about morality are important because they constitute the political values upon which theorists can plausibly claim reasonable people will be motivated by their sense of justice to act. Having said this, it may often be difficult to identify and categorise people’s views as ‘a view about justice/morality’.

The problem with precisely identifying the content of people’s views suitably categorised reflects a wider epistemic challenge. Elster describes the task of identifying people’s perceptions of justice as the descriptive study of justice (1995, p. 81). Such studies can be directed either at behaviour or at attitudes, and take place either in real-life contexts or in artificial settings such as surveys or experiments. These variables yield four options of approaches that theorists
might take: studying attitudes as manifested in verbal behaviour; surveys or experiments to study attitudes; experiments to discover behaviours; and finally by observing instances of what Elster calls ‘local justice’ (processes of wage formation, income transfers, allocation of goods other than money, and the like) (ibid., pp. 83-4). It’s not immediately clear that RU theorists can meet the demands of their methodology more satisfactorily with one of these approaches over alternatives. Furthermore, it’s not clear that any of them are sufficient to produce the kind of information that RU theorists are really interested in. There is a view that ‘empirical studies of justice are of little value in getting at what people really think about this subject, because they are either carried out in artificial situations (laboratory experiments) or else are hopelessly superficial (survey research’) (Miller, 2002, pp. 555-6). Given these problems, ‘empirical research may simply turn up a distorted set of ideas, biased by individual or class interest, cognitive failures of one kind or another, etc.’ (ibid., p. 556). Perhaps, then, rather than taking the conclusion that people’s views matter as a starting point and trying to develop substance from this empirical data, it is worth considering whether the methodological commitments that tie RU theorists to people’s views can guide the approach they should take.

The only requirement that I said demands RU theorists make their principles sensitive to people’s views is stability. This was because they necessarily inform the public political culture, compatibility with which is essential to establish stability for the right reasons. Actual people’s views about justice inform the PPC because it must at least partially reflect the shared values of a society. I said that there were numerous conceptions of the PPC and gave the following examples: an implicit or explicit function of the majority’s views (via a voting system); an overall set of values that makes the disparate views as consistent as possible (provided by social scientists); a model that aggregates the preferences of everyone in a society as well as possible (another task for social scientists); or a set of values that haven’t been vetoed by any groups in the society (again via a kind of voting system). Each conception of the PPC gives theorists a particular idea of the way in which the relevant data is collected. More work needs to be done to establish which empirical methods discussed in the
previous paragraph theorists should adopt to produce a plausible account of a society’s PPC.

The way that Rawls draws upon the values found in modern liberal democracies (freedom, equality, and fairness) makes it seem as though establishing the political values of a particular society may not initiate the kind of methodological concerns I’ve been discussing about how to collect the relevant data. He says that these values can be found in ‘the culture of daily life, of its many associations: churches and universities, learned and scientific societies, and clubs and teams’ and that that these and other main institutions are a fund of implicitly shared ideas and principles that are ‘at least familiar and intelligible to the educated common sense of citizens generally’ (2005, p. 14). It seems, then, that the relevant political values are already known and available to members of those societies. Given that Rawls was a member of the kind of modern liberal society to which he intended Justice as Fairness to apply, he was already aware of the values necessary to secure its stability. We might even think that ‘we, Western liberals, have such a thick commitment to mutual justifiability to rational persons [Valentini’s interpretation of the substantive essence of liberal justice] that it would be impossible for us to theorise about justice prescinding from that commitment’ (Valentini, 2013, p. 198). If this is the case, then the requirement to acknowledge sensitivity to people's views won’t have a substantive impact on the way that RU theorists go about their theorising (provided they are designing principles for the same public culture as the one they are within and familiar with). The impact of my conclusion may then be found in the idea that a theorist’s present context defines the scope of RU principles she may develop unless she finds a suitable approach to collecting the relevant empirical data of others (this in itself is quite substantial). If theorists want their principles to govern a different context to the one they are currently in, then the empirical issues discussed (of how to collect such information) needs to be addressed. They must also consider and address the interesting epistemic question regarding whether theorists can ‘know’ the content of the values of a culture of which they are not a part in any meaningful sense.
At the end of the final chapter, I suggested that democratic procedures are one way in which a theory may incorporate people’s views. Although this is at first a plausible suggestion, there is a huge literature surrounding the question of whether democratic procedures can accurately reflect the will of the people. Many suggest that it’s the voting system, rather than people’s views, that ends up being decisive (Riker, 1982). This is problematic because it means that the result is not fully informed by people’s beliefs. If this is the case then including a democratic procedure into the RU theory may not allow it to satisfy the requirements that people’s views inform the content of justice. In this case, theorists might have (legitimacy-based) reasons to adopt a democratic system, but doing so won’t allow them to simultaneously meet the requirement that their theory be sensitive to people’s views. Theorists must then reconsider the preceding questions of how to collect the relevant information.

The purpose of the discussion here has been to draw attention to the questions raised by my conclusions, and demonstrate the need for more careful attention to these empirical questions if RU theorists are to proceed by including people’s views into their deliberations.

If a theorist sits down tomorrow and acknowledges that she shares what I take to be the core commitments of the RU approach, then how will what I have said in this thesis affect what she does next? If she accepts the idea that people’s views are important to justice for more than stability reasons alone (or if she intends her principles to govern a context other than her own), then she will begin her task in a very different way; she will need to think about how to uncover the values, practices, and views of those whom she is interested in prescribing principles for. Although she is still committed to theorising for a utopia, uncovering and thinking carefully about these values gives her substance to work with and fulfil her simultaneous commitment to realism. If she rejects the second conclusion of the thesis and thus the significance of people’s views, then her task is still more burdensome as a result of the three more stringent commitments I’ve defended. They are, however, a good kind of burden, since they yield the best version of principles to govern a hypothetical but practically possible world, as I have sought to show.
Bibliography


Dworkin and His Critics: With Replies by Dworkin (pp. 3-29). Oxford: Blackwell.


Cambridge: Harvard University Press.


Princeton University Press.


844-864.


---. (2007). The Burdens of Public Justification: Constructivism, 
Contractualism, and Publicity. Politics Philosophy Economics, 6(1), 5-43.

Christiano, & J. Christman (Eds.), Contemporary Debates in Political 
Philosophy (pp. 41-60). Oxford: Wiley-Blackwell.


