The Roles of Moral Psychology in the Philosophy of John Rawls

Ross Hetherington

BSc MLitt

Submitted in fulfilment of the requirements for the Degree of Ph.D.

Philosophy

School of Humanities

University of Glasgow

April 2012

Abstract

This thesis explicates and critically considers the various roles played by moral psychology within the work of John Rawls throughout his career. In the second half of the 20th Century, Rawls's development of a sophisticated theory of justice in the social contract tradition played a significant part in reviving the study of normative political philosophy in the western world. Rawls argued that any theory of justice must be closely integrated with our best contemporary understanding of human psychology. Moral psychology is hence widely recognised to play an important role in Rawls's overall theory. But the precise role played has not been adequately examined. In this thesis, I identify six roles which moral psychology plays within the structure of Rawls's theory. Moral psychology must defend the idea that the model for a just society which Rawls proposes is realisable and stable (role #1). Moral psychology is also employed to explain how persons now have acquired what sense of justice they have (role #2). By showing that Rawls's just society can be realised and is stable, moral psychology is then subsequently used in the justification of Rawls's theory of justice – first by showing that such a society is not futile (role #3), and second by showing that the society is comparatively more stable than leading rivals (role #4). The account of the psychological capacities of the moral person is used to place the limit on the scope of justice (role #5). And moral psychological facts are also likely to be, in some sense, constitutive of the nature of morality for Rawls (role #6). These roles are discussed throughout various chapters. What alterations occur to the overall place of moral psychology following Rawls's later embrace of political liberalism is also discussed. The overall aim of the thesis is to produce an accurate exegesis on these matters, and in doing so indicate just how important moral psychology is within Rawls's theory, but also to indicate, clearly and starkly, just how much more psychological and sociological investigation needs to be done in if the theory is to be substantiated, given Rawls's own criteria.

To

Granma and Auntie Millie,

Nanny and Granda,

Auntie Marie,

and

Auntie Kathleen

Table of Contents

Preface

In the course of writing this thesis I have had a lot of help from the most wonderful people. I'd like to thank everyone who has helped me out over the years, but unfortunately, I'm afraid I will probably forget some of you. If we've spoken about my work – on this thesis and in philosophy generally – over the last few years, but you don't appear here, then please email me, and I'll try to get you added in. If not, please understand that you should be here.

Except in a few cases, these thank-yous are not arranged in any particular order. I mean to thank, for the most part, everyone as much as everyone else – the deeper thank-yous will be obvious, I hope. I'm cutting no corners in what follows, and I personally would be being false if I didn't try to convey here something of how I feel about you all. If you are not a fan of sentimentality, then please do feel free to skip over this (except for my examiners – please do read my thank-you to G.A. Cohen on p. xv).

I must first thank my Mam and Dad – Rob and Barbara Hetherington. I did not manage to procure any funding for this PhD. Their generous financial support has allowed me to study full time rather than part time.

But much more importantly, along with my sister Becky, they have also been wonderfully supportive personally. In no way could I have got through this without them; I can't imagine having a better family.

I next thank my long-suffering supervisor, Dudley Knowles. When I started my supervision, I was a poor writer and an overly ambitious thinker. He has managed to cure me of the first problem, and has even managed to have some success with the second. He has also been a stalwart confident when I have gone through difficult times. We haven't always agreed, and he still has questions which have gone unanswered. I only wish I will be able to, one day.

My next thank you is a smaller one – to my second supervisor Alan Carter. For various reasons, we talked about my work little. But he presented me with a pressing question early in the development of my thesis. Answering this question was significant for the development of my position and for my eventual methodology. Alan has started a new life recently in St. Ives. Good luck to him.

I must also thank my internal examiner Ben Colburn. Ben has not seen any of my written work, due to the critical position he is required to take upon it. But his puzzlement about what exactly I was arguing for in one of the first talks of mine he saw, and further

puzzlement, always combined with enthusiastic support, at later ones, was something of a shot in the arm at a crucial stage in the thesis. I am in no doubt that the thesis has achieved what level of rigour it has due to his insightful and sincere criticism and example. Keep it up – give me hell!

Various other members of staff in the philosophy department have given me good advice over the years: advice which has built up. I acknowledge the best pieces of advice below, as well as I can remember them. Thanks to: Fiona MacPherson – you encouraged me to do the PhD when I was unsure, and reminded me, at a crucial moment many years later, that we cannot assume every problem can be solved; Nikk Effingham – you argued that your hand in front of you need not be a spatially extended. It was just the sort of thing I needed to hear argued at the time to get me to see just how much philosophy is about getting away from prejudice; Michael Brady – this thesis still has too many footnotes, but it would have had even more if I hadn't heard one of your broadsides against them. Plus you made a comment about public ownership recently which was heartfelt, and helped me to extricate my mind from some bad mental company; David Bain – many a time in the pub you played devil's advocate against the left (or that's how I read it). I learnt from it; Stephan Leuenberger – you organised my viva, and were always supportive; Martin Smith – I turned up to your epistemology course, and I am sorry I didn't stay for the whole thing. If I've picked up anything from your polite, insightful way with questions, I'm a lucky man; Chris Lindsay – Your down-to-earth dedication to teaching and to us graduate teaching assistants has helped me to love tutoring, plus you lent me some great Pere Ubu records; Gary Kemp – you have always conveyed to me just how important you view research into Rawls is. It kept me going, and maybe if I'd thought about it more, I could have kept my spirits up more through the tougher times; Adam Rieger – since taking up the head of department, and indeed before, it's been obvious you're on everyone's side; Alan Weir – Alan, if I hadn't taken your philosophy of language class, I think things may have turned out a lot worse. You have the most tremendous patience with other people's ideas, and I've tried to have that as well; Victoria Harrison – whenever I've talked to you about how my work is going, you've always said excellent things; Sue Lock – You always seemed to view my excessively wide reading as an asset. Thanks for the encouragement for me to be myself! Jake Chandler – I remember talking over formal decision theories with you. I was suspicious then, but it's good to know your enemy. And you yourself were the very opposite of one of those!; Paul Brownsey – Thanks for making tutoring as fun and rewarding as it was; Richard King – no one teaches patient exegesis like an ancient philosopher. I learned a lot; Anna Bergqvist – your time at Glasgow started too late for me

to benefit from your obvious talent, but hearing that other people found the last stretch of the PhD hard gave me the boost to get the job finally done. Richard Stalley – Richard, thanks so much for your little observations about Rawls over the years up until your retirement, and thanks ever so much for turning up to the political philosophy reading group since. I appreciate so much your persistent dragging me back down to earth to explain myself.

The secretarial staff at the philosophy department have also been so helpful over the years, both for getting things done, and having a chat about anything other than philosophy! So thanks Susan Howell, Anne Southall, and Jane Neil. You're all irreplaceable.

I believe that I have had some of the best fellow postgraduates here at Glasgow that I can possibly imagine. The egalitarian and supportive atmosphere, which I believe anyone will find here (if they take the time to look for it, even a little), I feel I have been benefitted by to a degree I couldn't begin to discern.

My first thank-you must undoubtedly go to Robert Cowan. We both did our conversion masters together. We started our PhD's together. You finished before me, but then, you always were the cleverer one. Thanks so much for innumerable conversations over the years – I only know so much as I do (particularly about metaethics) because of you. You also read over chapter 3 of this thesis – it meant a lot to me that you found it clear. These were the purple pasta days …

I also owe a great deal to Stuart Crutchfield. Your nose for bullshit helped to keep me on the philosophical straight and narrow. I also couldn't have got through without someone to talk Skronk with.

Graham Peebles deserves a special mention. We had many long conversations about philosophy and about politics. They rarely overlapped, but I got a lot out of them anyway.

I also need to thank Ioanna Patsiladou. You read over the introduction and first chapter of my thesis, and they are a lot neater because of it. You've also always been a sympathetic friend through the rough times – and hanging out with you has always cheered me up.

John Donaldson – I've never quite met anyone who argues in the way you do. I've learnt a lot about philosophy from seeing how you approach it.

Gareth Young I have now known for many a year. You're turning out to me a fine philosopher, and sitting opposite you for a good year or so has undoubted enhanced both my philosophy, and my knowledge of beer. You'll go far – get working!

Neil McDonnell will no doubt be one of the leading lights in metaphysics one of these days. But he also always believes in everyone else as well. You've got me to think again, and again, and again a great many times – I appreciate it.

Chris Yorke – you were there for some very tough times for me. I'm still in your debt; thanks.

Ariel Cecchi must surely be one of the nicest guys I've ever met. About a month in the department, and he offers to proof-read a chapter of my thesis, when it's not even in his field. Thanks a great deal, Ariel – never put out that human fire.

Carole Baillie – you've always had an ear when I've found the work (and my life!) tough, and you've always managed to remind me that to ultimately approach philosophy well, you've got to approach it slowly. I greatly appreciate your proof-reading of the 5th chapter – it improved it greatly. And you yourself will do great whatever you do.

Giovanni Gellera has encouraged me in thinking that you can do good philosophy without skimping on the exegesis. Thanks for helping me to have a bit more faith in myself and my way of doing things.

Umut Baysan – you obviously want to get to the bottom of things. As do I! You're approaching your PhD in a perfect way, and I've appreciated your judicious comments on just about everything.

Akiko Frischutt and I are both people who find it difficult to stop before every question has been answered. This is a dangerous habit in a philosopher, but it's also a good one, and it is good to know someone who shares it.

Andy MacGregor has some crackpot views – just like all the great philosophers! I have always appreciated talking to someone who is willing to go out on a limb. I must confess: I always wonder whether I'm half-convinced.

Gavin Thompson as moved on from philosophy. But it's always great to see you, and you've given me some good things to think about over the years.

Stephanie Rennick – another person who kindly agreed to read a chapter of my thesis having only known me a few months. It was greatly appreciated. You'll be another one who will go far.

Renee Bleau – you've asked me some good questions over the last few years – one's which take some guts to ask. As you've learnt more philosophy, they've become more focused. But never stop asking the hard questions – to me or anyone else.

Chris Reid is certainly not an analytical philosopher. But he's not a continental one either. It's been good to bounce some audacious ideas of each other's heads over the years, Chris. You have an eye for deep issues.

Ben Wilson – though you decided philosophy wasn't for you, I remember some great conversations about free will and society, which I certainly took something from.

Beth Kahn – your enthusiasm was infectious, and you're most certainly always on the side of the angels.

Pat McDevitt, Alan Wilson – it was a pleasure reading through Raz with you, and facing all those stubborn questions as I tried to charitably read him each week. I learnt a lot from you guys, and I'm sure you will both go far.

There have been many other people come through the doors of Glasgow and sit its two excellent masters courses. I've enjoyed the talks I've heard over the years – some of them immensely – and a great many more conversations with many intelligent and likeable people. So for any former masters students whose names I have not reproduced here – thanks for your contribution to the department, and I hope things went well for you all.

I have attended only a small number of conferences over the years. I always left them thinking "I've got to stay in touch with all these guys." In some cases, I have even managed it. I would first like to thank Mar Cabezas and Carmen Velayos, and Chris Mills and Joe Horton, who organised conferences at The University of Salamanca (2010), on moral philosophy and the emotions, and The University of Manchester (2011), on political philosophy, respectively, at which I presented talks. Regrettably, I have not found space for the material from these talks in this thesis, but reflecting on the issues has no doubt helped it. The conferences were also superb, guys.

At said conferences, I remember enlightening conversations (at Salamanca) with the Late Peter Goldie, Chloë Fitzgerald, Melissa Stobie, Fabrice Teroni, Ulla Schmid, Stéphane Lemaire, Axel Seeman and Jesse Prinz, amongst others, and (at Manchester) with Andrea Sangiovanni, Amanda Cawston, Angie Pepper, Dean Redfern, Elizabeth Ellis, Felix Gerlsback, Garvan Walshe, Kimberley Brownlee (who I also spoke to about my thesis when she gave a paper at Glasgow recently – thanks for listening!), Liam Shields, Sabine Hohl, Sam Kukathas, Stephen Hood, and Yann Allard-Tremblay, amongst others. I can't remember talking any philosophy with him, but I also had few pints and a nice breakfast with John Wright.

I have also attended other conferences. Some people I met multiple times – I won't repeat their names, I afraid. At an excellent workshop on Motivation and Global Justice in York, organised by Kerri Woods, I talked to Carol Gould, Lea Ypi, Katrin Flikschuh, Simon Hope, Sue Mendus, Alex Bavister-Gould, and Martin O'Neill, again amongst others. Simon I had already met at the Stirling political philosophy seminar, and at the Glasgow undergraduate philosophy society – I found the talk he delivered at Stirling on the

circumstances of justice very useful in the early stages of my PhD, and I was happy to find reason to cite it again here. Martin O'Neill I also remember was particularly interested in my project – thanks for that.

I attended the festschrift for Hillel Steiner in 2009. I had good, though often brief, conversations there with Ian Carter, Jonathan Wolff, Eric Mack, Jethro Butler, and David Rhys Birks, amongst others.

Also in 2009, I attended a conference in Manchester on the political philosophy of T.M. Scanlon. I got a lot out of talking to Waheed Hussain, Michael Otsuka, T.M. Scanlon, and a great many others this time (my memory is unfortunately really failing me here).

I attended a conference on constructivism, in the same year I think, at which I talked to Andrew Williams about all that stuff – once again useful.

Finally, right at the beginning of my PhD, I attended the annual Law and Philosophy conference at the University of Stirling, which was organised by Ambrose Lee and Piero Moraro. I remember good conversations with Alice Walla, Antony Duff, Daniele Mezzadri, Jesse Tomalty, John Horton, Katherine Brooks, Kent Hurtig, James Dempsey, Massimo Renzo, Raymond Critch, Rowan Cruft, Matt Matravers and Sven Braspenning.

There are some acknowledgements which I would like to make to books I feel I have got a lot out of during the course of my PhD, but which I do not cite, or do not extensively cite, within the following thesis. I feel these volumes have shaped my thoughts and the way I go about philosophy as much as any conversations I have engaged in, or reflections I have had. So it seems fitting that they, and their authors should appear here. Hence I would like to thank Hillel Steiner for his *An Essay on Rights*, The Late Susan Hurley for her *Natural Reasons*, The Late Richard Wollheim for his *On the Emotions*, and The Late G.A. Cohen for his *Rescuing Justice and Equality*.

I am especially sad that I have not made more room to discuss Cohen's book. I had at one stage planned an entire chapter on it, but dealing with its arguments in depth would have simply taken away too much space which needed to go on more essential discussions (given the overall focus of the thesis – see the introduction). I also had the pleasure of meeting Cohen a few months before his untimely death – sat opposite each other on the meal at the second night of the Scanlon conference mentioned above. To my perception at least, we got along straight away, and I was extremely sad to hear that he had died.

There are some final random names I would like to mention, before putting a lid on all this effusiveness (no doubt some readers are rolling their eyes – I'm sorry but I make no apologies).

Dagmar Wilhelm was a masters student here at Glasgow before my time. She pops back every now and again, and we always have a good chat. A very daring thinker.

Paul Smith I know from Glasgow. Paul, you're one of the most honest people I know, and you've often made me think twice about things, so thanks.

Sarah Honeychurch was completing her PhD when I arrived. She attended my political philosophy reading group for years, put up with my overly pedantic reading manner, and in general taught me a lot and usually cut through the crap. Thanks for numerous chats over the years.

Glen Pettigrove, Pekka Vayrynen, Claire Batty, Robin Le Poidevin, Thom Brooks, and Claire Chambers all visited the department, either to give a talk, attend a conference, or even stay for a few months. I remember talking about my thesis and doing my PhD with all of them. I'm very grateful – particularly to Glen. I feel there must be many more visiting academics I have talked to over the years, but I cannot recall them all now. Whoever you are, you all have my thanks.

While I still lived in Edinburgh, prior to my masters, I was encouraged in philosophy by three great friends. Please step forward Dylan Wade, Phil Harris and Andreas Paraskevaides. Thanks for all your encouragement, and for telling me what to look out for.

James Dowey I also know from Edinburgh. He has always been interested in my work, and always reminds me of the Keynesian saying that common sense now is often what a theorist first hit upon centuries ago. I wish I saw him, and all my other former flatmates in Edinburgh, more often.

Andrew Wade I know from back home. Having recently come to the city, he, graciously and of his own free will(!), attended one of my final talks. I found the perspective from a brazen non-philosopher refreshing, and I still haven't worked out how to properly reply. The same goes for all my other friends from Bishop Auckland who have expressed their puzzlement at my studies. I always counted what any of you said as as important as anything I got out a book.

Andrew Holden is responsible for me getting into philosophy. For it was he who bought me that fateful copy of Bertie Russell's *The History of Western Philosophy* for my birthday in the third year of my original degree (which was in geography). I never even knew a subject like this existed, and eventually, I wanted to pursue nothing else. He is one of my oldest friends, and my life would have been much different without him around.

Finally: Cora, you've been a true friend through everything, and I can't thank you enough.

All of these people have helped make this thesis what it is, and it would be a much poorer specimen without them. But the thesis is not dedicated to them, but to various relatives, both living and dead, who have supported me throughout. They are Eleanor Marianne Hetherington, Amelia Ade Whittaker, Jean McCombie, William Gordon McCombie, Marie Longstaff, and Kathleen Winn. I'm only as good a person as I am because I was brought up with all of you around. To an extent I will never know, I owe to you what moral sense I have.

Addendum: I would further like to thank Jonathan Wolff for his diligent and good-humoured examination of this thesis.

Abbreviations of Rawls's Works

I use the following abbreviations for Rawls's books throughout the thesis. For full references, please see bibliography.

TJ	*A Theory of Justice*
PL	*Political Liberalism*
CP	*Collected Papers*
LP	*Law of Peoples*
LHMP	*Lectures on the History of Moral Philosophy*
JF	*Justice as Fairness: A Restatement*
LHPP	*Lectures on the History of Political Philosophy*

References are given in the format *LHMP*, p. 100, except in the case of *A Theory of Justice*, which are given in the format *TJ*, p. 477/418 or *TJ*, pp. 50—51/44—45 where the first number(s) refers to the page number(s) in the original edition, and the second to the respective page(s) in the revised edition. When the passage is absent in the revised edition, a second number will be absent, and vice versa.

Introduction

> The part of the book I always liked the best was the third, on moral psychology.
>
> John Rawls

John Rawls was arguably the greatest political philosopher of the 20th century, and a daunting figure even in moral philosophy. He believed that both subjects could only progress by the development of systematic and integrated theories, and the breadth and depth of his work is a testament to his pursuit of this conviction. Rawls's own theory contains numerous separate elements. Designed to fit together as a whole, a marked number of them have nevertheless been individually influential. This thesis focuses on the major element of the theory which has perhaps received the least attention: Rawls's moral psychology.

Was this comparative neglect warranted? In a word, no. First, it is obvious that Rawls thought the topic of moral psychology was important. The amount of attention he gave to it is enough to say this. The majority of the third part of *A Theory of Justice* – his key work – is concerned with moral psychology and related issues. In the passage containing the line I opened with, Rawls tells us that moral psychology was the area of his work that he most wanted to develop after the publication of *Theory*, but that replying conscientiously to his many critics eventually took him down a different path.[1]

Second, though Rawls's work on moral psychology has not been extensively commented on, I feel it has had a wider influence than has been recognised. In the work of Rawls's many students, what is going on is often illuminated by considering Rawls's work, frequently through his reading of the great historical philosophers. And often the influence is on issues within moral psychology.[2] Furthermore, debates in contemporary philosophical moral psychology have often taken Rawls as their point of departure.[3]

Third, and most importantly, what Rawls has to say about moral psychology is important, and one can learn generally from his approach. In particular, he has a very clear sense of what roles moral psychology should play in moral philosophy more generally. This can be found in the very way he structures his moral theory — Justice as Fairness.

---

1 This passage is from a set of unpublished remarks "My Teaching". See Freeman (2007a), p.6—7

2 The students in question include Christine Korsgaard, Sibyl Schwarzenbach, Thomas Scanlon, and Henry Richardson

3 See, for instance, important work by Thomas on self-respect, (1977—78), (1978), (1995), and by Deigh (1982), (1983) and Taylor (1985) on guilt and shame.

This thesis focuses for the most part on this latter aspect of Rawls's moral psychology: the various roles moral psychology 'plays' within Justice as Fairness. I see there being three sides to the study of moral psychology in Rawls. One is the study of the content of the psychology – the content of the actual psychological claims made by Rawls in describing his picture of the society of Justice as Fairness. Another is the study of the relevance of psychology within Rawls's more general methodology in moral philosophy, which incorporates his well-known method of reflective equilibrium. This has been recently elaborated on in depth by John Mikhail, going on to form the basis of an contemporary research programme linking cognitive science, psychology and philosophy.[4]

The third is the study of the roles that moral psychology plays within Rawls's normative theory — in particular in the setting up of the original position, and the argument from within it. It is this aspect of moral psychology within Rawls which my thesis focuses on. If my overall conception of the roles of moral psychology within Rawls's theory is sound, then I believe this work can help to properly orientate the study of the moral psychological content of Justice as Fairness. We will be able to appreciate its full richness, more precisely identify what problems exist for it, and which of these problems represent wider problems for the theory. It may also, in addition, contribute to the research programme instigated by Mikhail.

Before starting on the discussion of Rawls on moral psychology in chapter 1, there are a number of preliminary matters which are best considered in this introduction. At the end, there will be a brief summary of the coming chapters.

1. What is moral psychology? And how does Rawls understand this term? Roughly, we can say that moral psychology is the study of the thoughts and behaviour of human beings which make them moral beings. What these aspects are, how they are related to and interact with one another, and how they relate and interact with the other aspects of ourselves I take to be the fundamental issues in the field. I aim for this characterisation to be extremely capacious. Notice, then, that none of what I have said commits me to any particular views about the *structure* of the aspects of ourselves which make us moral creatures. The moral aspects of ourselves are not presupposed to be unified in any particular way, or to any particular degree. The idea that the moral aspects of ourselves are unified reaches its apogee with the traditional idea of the *moral sense*: a discrete moral module in the mind, often claimed to be found in the writings of the sentimentalist tradition in moral philosophy. This idea is most likely a psychological fiction, but I do not

4 See Mikhail (2011)

aim to enter into the debate at this level.[5] However, for convenience I do want a general term to cover the aspects of human beings in virtue of which they are moral creatures. *Moral sensibility* seems to me the best, suggesting as it does a general care, concern and responsiveness to moral matters in both thought and action.[6] Moral sensibility I shall understand to be something which is realised in ourselves. When taking about those aspects of us which make us moral beings both realised and nascent, I shall use moral psychology. These distinctions will be elaborated and reiterated in later chapters.

If that is what moral psychology is, how do we study it? There are two broad disciplines that attempt this: philosophical moral psychology, and empirical moral psychology. Philosophical moral psychology focuses on conceptual analysis and philosophical abstraction. It often relies to a great degree on introspection and intuition, and often makes use of literary examples or the invention of picturesque scenarios. It is (usually) concerned to elucidate and defend folk-understandings of moral psychology. Empirical moral psychology is a branch of psychology as an academic discipline. It focuses on quantifiable experimental results, and it often has less time for introspection and folk-concepts. These very broad characterisations must be understood to be caricatures. They are increasingly out of date. Recent years have seen both moral psychologists and philosophers paying more attention to each other than ever before, combined with an explosion of genuine cross-disciplinary work.[7]

Some much for a sketch of moral psychology as it stands. How does Rawls stand as regards to it? Rawls was pioneering in seeing the value and importance of modern empirical psychological research to the moral philosopher, though the material he relied upon was restricted (unavoidably, given his wider ambitions and commitments). Aspects of his own moral philosophy itself have in turn been influential on many research programs in the contemporary field.[8] However, Rawls's approach to moral psychology is still largely in the manner of a philosopher. One notable aspect, diverging from at least some contemporary philosophical work, is Rawls's breadth and systematicity. As with all his work, with Rawls's moral psychology you get a full package. You won't get engagement with precise theoretical debates, but everything will be covered in some form. In this, I see Rawls as closer to the great historical philosophers, and their approach to

5 On questioning the idea of a unified moral sense, see Flanagan (1991), pp.266—267. For recent developments, see Cushman, Young, and Greene (2010), but also see Mikhail (2011)

6 My way of setting up things here has benefited from Wren (1991), esp. chapter 1.

7 For a sample through the past 20 years, see May, Friedman, and Clark (eds.) (1996), Sinnott-Armstrong (ed.) (2008), and Doris and The Moral Psychology Research Group (2010)

8 See, for example, Gibbard (1982), Mikhail (2011), Hauser, Young and Cushman (2008), and Roedder and Harman (2010).

moral psychology, than many contemporaries.

In this thesis I shall not make extensive use of the contemporary empirical literature in my assessment. The element of Rawls's theory on which I am focused does not call for this. I believe that we need an outline of the roles that moral psychology plays within the structure of Rawls's theory prior to engagement with its content, and it is in assessing the content that empirical data would have obvious importance. I also believe that philosophical and exegetical engagement must come prior to thorough empirical engagement in order for a philosopher to do the best by both disciplines (though not that they can ultimately do without it). Hence, in view of its principle subject matter, and my own current expertise, I have put aside empirical reports and studies for this thesis.

2. Rawls's thought developed greatly over his career. Did his moral psychology change with it? Given the controversy over the differences, or lack of them, between the earlier and later Rawls, it seems wise for me to say something at this outset.[9]

I believe that neither the substance nor the roles of moral psychology undergo any fundamental alterations in the course of Rawls's career. There's some change, but the essentials display a great deal of continuity: from the earliest presentation in the article, "The Sense of Justice"[10], through to the elaborations in *A Theory of Justice*, then *Political Liberalism*, and beyond. What changes there are I shall indicate at relevant times throughout the thesis.

However, as everyone knows, there is a fundamental change between the position Rawls advances in *Theory* and that advanced in *Political Liberalism*. The theory of justice defended in the original book is put forward as a comprehensive moral theory, or at least the kernel of one.[11] In the later book, this theory is transformed into a specifically political theory.[12] I need a methodology for coping with this change throughout the thesis.

The one I propose is this. For the first four chapters, I bracket the material which is specifically used to develop the idea of a political conception of justice. But I otherwise make use of material from the later works. Not all this later material demands to be kept and treated strictly separately, because not everything in Rawls's later work stems from trying to describe the idea of a politically liberal regime. For each addition or modification, we can ask "Can I imagine Rawls introducing this new material even if he *hadn't* embraced political liberalism?" As it turns out, the answer is "yes" for most of it. In the fifth chapter,

9 Examples of articles that attempt to defend the outline the continuity of Rawls's thought include Wenar (2005) and Estlund (1996). Articles which make the case against include Barry (1995).
10 *CP*, chapter 5
11 *PL* p.xvii
12 See *PL*, pp.12–13

I then discuss the idea of a political conception of justice. I attempt to see just what implications this has for the preceding discussions of moral psychology. The sixth chapter then follows. The position I wish to defend there is best discussed presupposing the introduction of the idea of a political conception to the subject of Rawls's moral psychology, and in addition the focus of the chapter represents an appropriate concluding topic.

As well as making use of material from the later works, I will also make use of Rawls's lectures on the history of moral philosophy and political philosophy. The same exegetical approach seems warranted. The ideas of the past philosophical greats were readily incorporated by Rawls into his own philosophy.[13] It would be overly cautious to discount such a resource. I should note that I shall primarily attend only to Rawls's own lectures on these writers, and not the writers themselves. Whether Rawls is correct or incorrect in his examinations is not of central importance. We are concerned with the ways in which his own understanding of these authors' doctrines might help us to understand his theory.

The necessary preamble is out of the way. Below follows a summary of the coming chapters. Each of these chapters is composed of one or more numbered sections. Many of these sections are composed of further subsections, and a few of these subsections have further subsections themselves. Sections will be indicated throughout the text in the following way: section 5. Subsections will be indicated throughout the text in the following way: subsection 5.2. The more traditional section symbol, §, will be reserved for referring to Rawls's work, as he uses it extensively.

Chapter 1: The Roles of Moral Psychology: This chapter will introduce the position and roles that moral psychology plays within Rawls's theory, elaborating on why Rawls introduces a moral psychology at all, and why in such depth.

Chapter 2: Moral Psychology and Justification: This chapter examines the role that moral psychology plays in the justification of moral principles from the original position. More specifically, it investigates an ambiguity in Rawls's account of this role, which may have significance for the outcome of the argument.

Chapter 3: Moral Psychology as Constitutive: This addresses whether, as some writers

13 See Samuel Freeman's foreword to *LHMP*, p.xi—xix.

have claimed, moral psychology plays some kind of foundational or constitutive role in Rawls's theory. Addressing two of these writers, I argue that it does not play a foundational role. I agree it does play a constitutive role, but not in the ways that some have claimed.

Chapter 4: The Conception of the Moral Person and Moral Psychology: This chapter presents, in its minimal details, the character of the moral person in Rawls's theory. This represents the basic starting point for any assessment of Rawls's wider psychological claims.

Chapter 5: Moral Psychology in Political Liberalism: Rawls's transformation of his theory from a comprehensive to a politically liberal one might be thought to have followed from problems with aspects of his moral psychology. It might be thought to lead to alterations in how his moral psychology is to be conceived. My aim here is figure out how to assess these claims. To do this, I reconstruct Rawls's reasons for revising his theory, and observe how moral psychology within his theory subsequently fares.

Chapter 6: The Scope of Justice and Moral Psychology: This chapter analyses Rawls's various accounts of the scope of justice, and defends one of these accounts against the others as most morally defensible, assuming a contractualist theory, and as also the most fitting with his psychology. The end of the chapter then highlights further problems which nevertheless remain with Rawls’s position.

Chapter 1: The Roles of Moral Psychology

This first chapter proceeds as follows. In section 1, I look back to Rawls's introduction of a moral psychology into his theory in the earliest articles presenting Justice as Fairness. Section 2 then presents Rawls's theory more generally, as it was eventually developed. Following from this, Section 3 describes the different, overlapping roles that moral psychology plays in Rawls's theory.

Section 1: Moral Psychology in the Early Rawls

Imagine you think you know the requirements of morality. Now imagine you know what would have to be the case, psychologically, for people to act in accordance with the requirements of morality. Morality says: you should act this way. Moral psychology says: people can act that way. But suppose not everyone agrees that what you think are the requirements of morality *are* the requirements of morality. How are you going to decide if you're right or they're right? So imagine you hit upon this: to develop an account of what justifies your requirements of morality, rather than the others. This account may also include psychological statements. Put together these two normative elements, and two psychological elements, and you will have something you might want to call a *theory* of morality. Maybe you'll want to add to it later, but for now let's just leave it be.

A remarkable element of Rawls's earliest formulation of Justice as Fairness[14] – his moral theory – is that the psychological element which corresponds to the justificatory aspect of the theory is in important ways distinct from the psychological element which corresponds to the requirements of morality. The psychology which is appealed to in the justificatory aspect of the theory is *not* put forward as a moral psychology. It need not even be put forward as a genuine theory of human psychology at all. Nevertheless, it plays a key role in the early justification of Justice as Fairness. Once this role is completed, however, we are left with the question of whether *we ourselves* – normal human beings – can be moved by the requirements of morality which have been defended. For this, we need a separate account of moral psychology.

Justice as Fairness, from its earliest presentations, included a moral psychology. The early articles I am about to discuss are "Justice as Fairness", and "The Sense of

14 I write Justice as Fairness as a proper name (capitalised) throughout the thesis. Note this is not Rawls's practice, and I have not altered quotations by him.

Justice".[15] I begin with the general argument found within these early papers in order to present the introduction of a moral psychology into Rawls's philosophy in its earliest and, we might expect, simplest form. Observing the structure of his account at this early stage should help to get a clear view of why it was necessary for him to produce a moral psychology, and of the particular issues it was designed to address. From such a starting point, we should also be able to pick up on whatever additions and alterations he later made to his account.

The best way to understand the overall argument in "Justice as Fairness" is to see that Rawls's primary aim is to point out the deficiencies of the conception of justice found in classical utilitarianism, whatever that theory's other virtues. Utilitarianism

> assimilates justice to benevolence and the latter in turn to the most efficient design of institutions to promote the general welfare. Justice is a kind [read:variety] of efficiency [which is applicable given certain conditions].[16]

Elsewhere, Rawls puts forward his earliest statement of his two principles of justice in this article, but his general approach does not depend on these being precisely correct. As he makes clear, they simply need to be representative of a certain family of principles which acceptably represent individuals' freedom and equality within shared institutions.[17] Now, assuming certain circumstances, institutions embodying such principles may be able to be derived from the principle of utility. This was the approach of liberal utilitarians, such as Mill.[18] But Rawls proposes a different derivation of the principles: one which procures them more directly, and which holds out better hope of explaining the importance we attach to justice,[19] and the force of the feelings associated with it,[20] without simply appealing to intuition.[21]

He asks us to consider what kind of principles mutually self-interested and rational persons, roughly equally situated within shared practices, would agree to in order to generally assess claims against those practices, knowing that they themselves must commit to any principles they propose and which are accepted.[22] Rawls's claim is that the

---

15 See *CP*, chapters 3 and 5
16 *CP*, p. 64. I have added the text in the square brackets, which I take to make clearer Rawls's meaning here. This shall be my standard practice throughout the thesis.
17 See *CP*, p. 48
18 See Mill (1863) chapter V
19 See *CP*, pp. 59, 67
20 *CP*, p. 68
21 *CP*, p. 52
22 See *CP*, pp. 52—55

principles we come up with through reflection on such a thought-experiment will to some degree correspond to the kinds of principles we intuitively think of as principles of justice. The requirements imposed in the hypothetical scenario on the self-interested and rational persons are those of fairness — hence we are able to account for the intuitively appealing idea that fairness is "the fundamental idea in the concept of justice".[23] Overall, this contractarian[24] conception of justice is thought to be superior to the utilitarian one from the perspective of supporting, explaining, and defending our everyday understanding of the importance of justice, and its association with fairness.[25]

However, the proposed hypothetical scenario only delivers us a derivation of the principles of justice for institutions.[26] The individuals within the scenario are purely self-interested, and it is stressed that their psychology is at best a truncated version of ours.[27] How actual persons will act when faced with the demands of the endorsed institutions in particular cases cannot to be derived solely from the features of the hypothetical contractors. In the original presentations of Justice as Fairness, the individuals within the scenario are only "required" to make a commitment in advance due to possessing roughly equal power and ability, and their being uncertain about what the future might bring. This situation forces restraint in the name of their own self-interest.[28] With such an origin, the commitment made cannot be expected to motivate such self-interested individuals on *all* occasions, particularly if it ever happens that rough equality no longer obtains. The expression of a general commitment to principles of justice does not imply a commitment to the requirements of those principles in particular circumstances. Rawls makes this very clear in "The Sense of Justice".[29]

Having derived the content of principles of justice for institutions by reference to the agreement of mutually constrained and self-interested agents, Rawls now has need for a separate account of how actual persons could come to be motivated by those principles directly in particular circumstances, in potential contradiction to their own self-interested desires. Hence the account of the moral psychology of the sense of justice: describing its

---

23 *CP*, pp. 47, 59

24 It has become common to distinguish between "Contractualism", meaning moral theories which make use of the notion of a social contract, but which place moral limitations on that contract, and "Contractarianism", meaning moral theories which make use of a social contract, but which do not place any moral limitations on the contract, and hence only embody prudential considerations. I have little use for this additional piece of jargon. Throughout the thesis, I use contractarian and contractualist interchangeably. See section 2 and subsection 15.1 for further elaboration on the idea of contractualism.

25 *CP*, pp. 71—72

26 *CP*, pp. 47—48, 63. See also pp. 99—100.

27 *CP*, pp. 56—57

28 See *CP*, pp. 53—54

29 *CP*, pp. 99—100. The idea is also presented, though less prominently, in "Justice as Fairness": see pp. 56—57, 61—63

development, its relation to other sentiments and attitudes, and the sense in which it expresses the principles of justice for individuals.[30] This developmental account is stipulated to be purely hypothetical – it may be that the precise development described would never occur. But something like it, and the relationship it suggests between our sense of justice and other sentiments, is taken to be plausible and compatible with the analysis of justice being presented.[31] From combining the account of the principles for institutions, the account of the sense of justice, and the account of the duty of fair play, the principles for individuals are obtained.[32]

I shall comment briefly on why Rawls appears to have taken the approach that he did. To defend his two principles of justice, he wanted to avoid appealing directly to our intuitions. This would be to fail to engage with utilitarian rivals on their own level: to do this requires that we develop some kind of deeper justificatory theory to explain and fit those intuitions within a broader system. More particularly, he wanted to be able to incorporate two key insights. One is that justice presupposes competing interests that people will be willing to press on one another, and which must be arbitrated.[33] The other is that people are motivated by considerations beyond mere personal advantage to settle such arbitrations — even in their own case, though admittedly often to a lessened degree.[34] One way to characterise his strategy is the following: presume everyone's self-interest first – which motivates the need for justice, after all – then place restrictions on such persons such that their institutions will be fair between such claims. In keeping the account of the virtue of justice out of the way at first, we make sure that we are addressing the central concern of justice, and not simply writing an edifying discourse on the just. Once the requirements of justice are set, we can then be sure to get an appropriate picture of the just person. It happens that, on this view, the sense of justice turns out to be something which almost everyone can be expected to possess to a sufficient level.[35] This is taken to be a serious advantage for the theory. I don't see how to be sure that these were the exact considerations which went through Rawls's mind, but they do seem to make good sense of the texts.

The need to accommodate the observation that justice concerns the arbitration of conflicting claimants, where neither is willing to back down through personal attachment, leads Rawls to derive the principles of just institutions by sole reference to self-interested agents facing each other within fair conditions. The need to account for our concern and

30 See *CP*, pp. 100—112
31 *CP*, pp. 100, 115
32 Putting together *CP*, pp. 59—63 and 112—116
33 *CP*, pp. 56—57
34 *CP*, pp. 62—63, 110—112
35 *CP*, pp. 112—113

attachment to justice itself is only then addressed, through proposing a moral psychology. The set up is fairly straight forward once it's understood, and why there is a separation between the derivation of principles and the account of the sense of justice is clear. Over the years, however, things were to become slightly more complicated...

Section 2: The Structure of Justice as Fairness

Justice as Fairness, in its full and final complexity, is more difficult to summarise. One way to review its structure is to proceed from the structure of *A Theory of Justice*, referring to discussions from Rawls's other works when necessary or helpful.[36]

Part One of *Theory* presents us with the statement of the principles of justice,[37] a specific group of arguments for them, and the arguments for the methodology which underpins the whole approach.[38] The *role* [39] (fair arbitration of claims within shared institutions — further elaborated in subsection 3.3)[40] and *subject* (the basic institutional structure of a single society)[41] of justice, and the circumstances which make it possible and necessary that justice obtain – the *circumstances of justice*[42] (see section 7) – are presented in order to set up the discussion, as they were in "Justice as Fairness". Intuitive considerations in favour of the principles of justice are first put forward.[43] But the main argument for the principles consists in deriving them from the original position.[44]

To develop the original position, *formal constraints* on the concept of right are first introduced. Rawls gives five such constraints: *universality*, *generality*, *publicity*, *ordering of claims*, and *finality*. It is inessential to discuss each of these now. Some of them I will return to. It suffices to say that they are all conditions described as intuitively morally reasonable to impose on *any* conception of justice — justice being just one virtue within The Right, or rightness, in general.[45] These formal constraints, however, do not themselves

36 Rawls summarises this structure for us at *TJ*, pp. 579–580/507–508. I am not alone in starting from this structural overview, and claim no originality for it: see Freeman (2003) pp. 279–280, (2007b) pp. 145—146

37 *TJ*, pp. 60—65/52—56

38 *TJ*, pp. 46—53/40—46

39 I introduce Rawls's terminology in italics throughout this section. Note that italicised words in other sections are not necessarily Rawls's terminology.

40 *TJ*, pp. 4—6/4—6

41 *TJ*, pp. 7—11/6—10

42 *TJ*, pp. 126–128/109—112. As will be outlined the section 7, this is not quite the right way to characterise Rawls's understanding of the circumstances of justice.

43 *TJ*, pp. 65—83/57—73. This is noted by Brian Barry (1989) pp. 213—234

44 *TJ*, chapter 3

45 See *TJ*, pp. 130–136/112–118. Note that these are not called formal constraints in that they follow logically or conceptually from the concept of right. Rawls states he wants to avoid that question. They are simply described as reasonable constraints.

serve to sufficiently narrow the range of principles we might adopt.

The idea of the *original position* is developed to generate the further constraints. The original position is Rawls's ultimate development of his idea of placing self-interested, rational choosers within a situation which forces them to conform to the constraints of fairness. In *Theory*, we are the choosers, as placed behind a *veil of ignorance*. A veil of ignorance conceals from us any knowledge of our eventual places in the subsequent society, or of our individual native and acquired abilities and propensities. We instead only know the general facts of human psychology, and that as a society we face the circumstances of justice.[46] In later publications, Rawls re-characterises the inhabitants of the original position so as to make them each a representative of a single *free, equal, rational and reasonable* person (see section 8) living in a just society.[47] The parties in the original position are no longer specified simply to be rational and self-interested. This would leave their possible interests undetermined. Instead, (see chapter 4) the parties are interested solely in protecting the interest that those they represent have in being free, equal, reasonable and rational persons.[48] This characterisation has the advantage of making it clear why the persons in the original position can be expected to be motivated only by self-interest. If we ourselves were to have a veil of ignorance cast over us, why would we be expected to be suddenly unmoved by our existing sense of justice?

The inhabitants of the original position are aware that any decision they come to must be able to be kept by those they represent. This is because the agreement is to be final, i.e. meet the finality requirement. The reason for this requirement is that the chosen principles of right are to govern the fundamental arrangements of the whole of society, and substantially determine the life-prospects for all who live within it.[49] Given the agreement is a one-off, there will be no reason to make an agreement that cannot be kept, as at least some of the interests which are meant to be protected by the agreement will not actually be protected, and will have no future chance of being protected. Because of this, the agreement made and the reasoning for it makes heavy reference to the facts of human nature (see subsections 4.3 and 6.1).[50]

Principles are then derived from considering the choices of the inhabitants of the original position. These principles are no longer simply the principles of justice for

---

46 *TJ*, pp. 136—142/118—123

47 See *PL*, pp. 24—25

48 See *CP*, p. 312, *PL*, pp. 73—74. The importance of this revision is stressed in the introduction to the revised edition of *Theory*: see p. xiii. See also *CP*, pp. 417—418.

49 *TJ*, p. 13/11—12

50 See *TJ*, pp. 137—138/119, 175—177/153—155

institutions, but also include the basic principles, duties and virtues for individuals.[51] In addition, the principles of right in general are also chosen in the original position.[52] I shall not elaborate all the various moral principles which Rawls derives from the original position.[53] But I shall put down the two principles of Justice for institutions which Rawls derives from the original position, for any moral or political philosophers who have been *incommunicado* since 1957[54] (I shall also have reason to refer to them in later discussions).

> a. Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value.
>
> b. Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices 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.[55]

Here also is the duty of justice which applies to all free and equal, rational and reasonable persons. It has two components.

> first, we are to comply with and to do our share in just institutions when they exist and apply to us; and second, we are to assist in the establishment of just arrangements when they do not exist, at least when this can be done with little cost to ourselves[56]

The original position is representative of the fact that Justice as Fairness is a variety of *contract theory*.[57] Contract theories attempt to lend justification to moral principles and precepts by showing how those moral principles would be those agreed to by agents

---

51 On duties, see *TJ*, pp. 108—117/93—101. On the virtues, see *TJ*, pp. 433—439
52 *TJ*, p. 333/293
53 See *TJ*, pp. 333—340/293—299, 342—350/301—308
54 "Justice as Fairness" was published in 1958
55 *PL*, pp. 5—6. See also *PL*, p. 291, *JF*, p. 42. Note these are revised statements of the basic liberties following *PL*, lecture VIII, which replied to criticisms in Hart (1975).
56 *TJ*, p. 334/293—294. See also pp. 115/99, 474/415
57 *TJ*, pp. 11—13/10–11, 15–16/14–15

situated with respect to each other in relations conducive to "informed, uncoerced"[58] and binding agreement. In Rawls's theory, the situation modelled is of free and equal rational agents being constrained by fair or reasonable conditions.[59] Rawls's hope is that the original position is the best *contractarian procedure* which can be used by a contract theory, and that, when it is specified correctly or adequately, it selects just one set of principles from those available, or at least indicates that one set has more going for it than the others.[60]

I myself distinguish between contract theories, and contractarian procedures. This is based on the thought that contract theories are moral theories which incorporate contractarian procedures as part of their structure, and employ them in the justification of moral principles. But in a contract theory, justification need not be conceived to proceed solely from a contractarian procedure. In Rawls it does not, as shall shortly be noted.

The use of a contract theory, and its particular specification, is based on the consideration of *fundamental ideas* which, on reflection, appear to underlie the political and social conflicts we currently acknowledge.[61] The theory developed, though it employs some highly abstract ideas, aims to engage with the real conflicts and problems actually faced.[62] Faced with social division and disagreement regarding what our values would have us do, Rawls proceeds on the assumptions that (1) our values might include shared values, and hence that the conflicts of value or interest in society may not stretch right down to the very bottom and (2) common ground may hence be able to be found, if we investigate carefully and sincerely.[63]

The fundamental ideas are *normative concepts and conceptions*. They tell us how we should be and should act, and reflecting on them is meant to guide our actions and correctly orientate our thinking.[64] For Rawls, a concept specifies "the meaning of a term," a

---

58 The phrase comes from Thomas Scanlon's work. See Scanlon (1998) p. 153. I take it that all contract theories can agree to this wording, differing in what they think people being informed and uncoerced requires.

59 *TJ*, pp. 12–13/11–12, 19/17. Fairness is the concept emphasised in Rawls earlier papers and *Theory*, reasonableness is emphasised in Rawls's later papers and books. The two concepts are distinct, but they are closely related. Reasonableness is discussed in depth in subsection 8.2

60 *TJ*, pp. 121–122/104–105

61 Rawls's fundamental ideas are set out in *TJ*, §1–4, but what they are, and what it even means to call something a fundamental idea, is presented much more explicitly in *JF*, pp. 1–2, 5–14, 18–26. See also *PL* pp. 4–5, 8–9, 15–22, 43–46. Rawls's later political liberalism requires further fundamental ideas. He specifies these at *PL*, p. xvi–xvii and 43

62 See *PL*, p. 43—46

63 The idea that the justification of moral and/or political theory, if possible, proceeds on the basis of shared values and background assumptions, admittedly which may have to be clarified and interpreted, is defended in *TJ* at pp. 580–583/508–511. The assumption is presented vividly at the very start of *CP*, chapter 16, pp. 304–305, *JF*, pp. 1–2. See also *PL*, p.43–46.

64 *PL*, pp. 8–9, 11–15, and also 43–46. For a particularly explicit presentation of Rawls's understanding of how political philosophy might guide our thought and action, see *JF*, pp. 1–6.

conception is a specification of a concept so as to include "the principles required to apply it," and "idea" is a general term covering both concepts and conceptions (see further subsection 3.3).[65] These ideas do not arise from nowhere. They are explications of values and commitments we are already taken to have, or at least which we can, through reasoning, brought to have through reflection.[66] Rawls's work is not addressed to those who do not or could not rationally come to recognise these values and commitments.

As Rawls's work progresses, he begins to talk less about theories of justice, and more frequently about conceptions of justice. I think that he views conception as a looser word than theory. The latter suggests a level of systematicity and rigour not required in examples of the former. I shall refer equally to Rawls's theory of justice and Rawls's conception of justice, choosing whichever word seems most fitting and clear at the time.

One central normative idea, which I shall mention again now, is a conception of the *person* (subsection 3.1) as a free, equal, rational and reasonable being.[67] In the earlier work, reasonableness is not mentioned when personhood is characterised, and people are only described as free, equal and rational.[68] But later comments make it clear that this concept was always present implicitly.[69]

Finally, all the various components of our theory so far assembled – our principles of justice, our formal constraints and the version of contract theory we derive the principles from, and the fundamental normative ideas that underwrite our contractarian device – should also be tested against the requirements of *the reflective equilibrium methodology*, often informally known as reflective equilibrium.[70] This is likewise true of the components yet to be put in place in Parts Two and Three of *Theory*.

I understand the method of reflective equilibrium roughly this way: to justify a moral theory, we should engage in the comparative examination of the various distinct moral theories and conceptions available to us, refining and developing them in order to render their differences vivid, and then should assess them against our considered moral judgements and attitudes, which can be similarly revised, to see whether any one theory wins out on due reflection. Upon reaching such a state, our moral judgements are in

65 See *PL*, p. 14 fn 15, and *TJ* p. 5/5. Fundamental ideas need not only be normative. Some, such as the first subject of justice being the basic structure, and the original position, are primarily or partially introduced for methodological or theoretical purposes. See *PL*, p. 14 fn 16.

66 *PL*, p. 45 and *TJ*, pp. 21–22/19, 587/514

67 See, for example, *PL*, pp. 18—19, 29—35, 48—54, *JF*, pp. 18—24

68 For example, *TJ*, pp. 252, 574/503

69 See, for example, *PL*, pp. 25 fn28, 53 fn7

70 *TJ*, pp. 46–48/41–42, plus also *CP*, pp. 286–289. I say informally, because strictly speaking reflective equilibrium is not the name for the whole of the methodology, but only for the end point the methodology aims for – one in which our theory and principles, and our considered judgements are in equilibrium, and in equilibrium due to our reflection.

equilibrium with each other and our chosen theory, and they are in equilibrium on the basis of *due reflection* (hence the name).[71] I comment very briefly on this methodology in subsections 3.5 and 5.1 below, but largely put the examination of it aside. A thorough description and critical assessment of this methodology would take a lot of space, and would be misplaced given the focus of this thesis.[72]

The initial stage of Rawls's theory, found in Part One of *A Theory of Justice*, lays out the claims and assumptions of that theory found at its most abstract levels: both normative claims and others. The later parts serve to introduce more and more concrete considerations. These are used to verify whether the theory fits with our considered judgements on due reflection, or else revises and extrapolates those judgements in acceptable ways.[73] I shall later try to elucidate the relationship between the fundamental ideas, and the development of the rest of the theory in terms of the notion of specification (subsection 5.2). Right now, I'll mention that the fundamental ideas should not be understood as *foundational* ideas, if this is taken to mean that the rest of the theory entirely rests on them as a foundation.[74] Rather they are simply the most abstract ideas within the theory – more particular ideas and more concrete data have an equally important place.[75] In saying this, I do not want to rule out the possibility that certain aspects of Rawls's theory are indeed foundational. It has been observed that it is unclear whether Rawls commits himself to a thorough-going coherentism, or whether what he says is compatible with some kind of moderate foundationalism.[76] I believe my thesis can say what it needs to say without resolving this issue

The elaboration of the theory from its most abstract elements onto concrete institutional, social and psychological conceptions is an essential part of the justification of the theory. Justification is said to stem from "everything fitting together into one coherent view" (see further subsection 3.5 below).[77] The abstract level of the theory possesses only provisional justification. The full justification is conditional upon the development and defence of a more concrete conception of the society which would enact the principles of justice, and a concrete conception of the psychology of the members of that society.

Part Two of *Theory* is concerned with interpreting how the principles of justice could be realised in institutional form. The aim of this part is to show that we can conceive

---

71 *TJ*, p. 20/18

72 Though I have substantial disagreements, the writer whose interpretation of Rawls on the method of reflective equilibrium is closest to my own is Scanlon (2003).

73 See *TJ*, pp. 579—580/507—508 plus also p. ix/xix, 95/81, 192/167—168, 195/171

74 For example, *JF*, p. 31.

75 For example, *PL*, p. 45

76 For this debate, see, for example, articles by DePaul (1986) and Ebertz (1993)

77 Ibid. p. 579/507. See also p. 21/19

of institutions which fit some acceptable interpretation of the principles. Key moral and political concepts found within the principles, such as 'basic liberties', remain too vague and ambiguous when it has not been specified how they might be instantiated in concrete institutions.[78] If we cannot articulate the relevant institutions, we may be forced to conclude that the principles are simply poorly formulated, or that the fundamental ideas underlying them are empty or inapplicable. At the end of Part Two, we have the bare but adequate bones of the institutional structure of the society of justice as fairness, including the rights and duties for individuals.[79]

But that we can conceive of a set of practices and institutions which match up to the principles and conceptions is still not sufficient for the full justification of the principles. To see this, we first need to distinguish between the description of just institutions, and the realisation of those institutions. In describing an institution we are describing "an abstract object," in other words "a possible form of conduct expressed by a system of rules."[80] The abstract object may or may not have a realisable counterpart. Describing a system of rules does not tell us whether and to what extent people can act in accordance with those rules. Hence, at this stage, we are only describing things normatively. When it comes to defending these institutions, Rawls requires that we do not rely only on normative assertions. We must also consider background empirical theories from the human sciences and humanities: in general, facts about human nature and psychology.[81] At certain points in Part Two, (and in Part One: see section 4) he does this quite explicitly.[82] But these scattered considerations and presumptions Rawls employs do not amount in themselves to a full moral psychology capable of defending the realisability of the society so far sketched. Instead, they presuppose one. We need to present a sufficiently complete account of the character of the people who would live their lives under such institutions. If it cannot be plausibly argued that human beings could maintain such institutions if they were set up,

78 On the topic of the basic liberties, see *TJ*, pp. 201—251/176—220. To see how seriously Rawls took the idea that moral conceptions must be able to give a viable institutional interpretation, observe his response to objections from Hart (1973) in *PL*, lec. VIII (see also *TJ*, p. /xii)

79 See *TJ*, pp. 114–117/98–99, 333/293, 337–340/297–299

80 *TJ*, p. 55/48

81 Psychology is just one particular discipline within the human sciences and the humanities, each having their particular domain. A theory of institutions needs to attend to not only psychology in this sense, but also sociology, history, political science, economics, geography and perhaps even human biology. Rawls indicates the relevance of most of them at various points in his theory. See, for example, references to the relevance of history (*TJ*, pp. 200/175–176, *PL* pp. 231–240), economics and political economy (*TJ*, pp. 258–259/228–229, 265–274/234–242), political science (*TJ* pp. 223–234/196–206) and human biology and evolutionary theory (*TJ*, pp. 502–504). This attitude fits with his general non–reductionist sympathies in moral theory. See *TJ*, pp. 577–578/506–507, and *PL*, pp. 86–88. For simplicity, I shall generally simply talk about psychological facts and theory. For the purposes of this thesis it is unnecessary to engage with the problems of the status of and relationships between the different human sciences and humanities.

82 See discussions on need for a legal system (p. 240/211), political economy (p. 260/230) and civil disobedience (p. 387/339–340).

then our theory remains unjustified overall.

Part Three of *Theory* takes up the task. The shape of the institutions required by justice is already in place. What needs to be defended is that, growing up and living under such institutions, people will be motivated and will act so as to sustain them. Their developed moral inclinations, in particular those associated with their reasonableness, one of the most important of which is the *sense of justice*,[83] must be strong enough to win out against any opposing motivations which would lead to the corruption of the justice of the institutions if unchecked.[84] If Rawls successfully argues for this, he will have argued that the institutions realising the principles of justice are sustainable, and hence that a just society meeting the criteria of the principles of justice is possible over time. The principles of justice argued for in Part One, and the theory in general, will then be justified.[85] The argument Rawls presents in Part Three to secure the justification of the principles of justice is commonly called the *stability argument*, or the *argument from stability*. I shall follow this convention.

The moral psychology found in *Theory* is similar in most respects to that found in "The Sense of Justice", and is maintained in roughly the same form throughout the rest of Rawls's career,[86] though the requirements of *Political Liberalism* do, as noted in the introduction, lead to some alterations. Alterations in both the role and content of the moral psychology, from the *earlier* philosophy to the *later*, will be addressed throughout subsequent chapters. I shall often distinguish between the comprehensively liberal and politically liberal periods of Rawls's work as earlier and later, except where otherwise indicated. I shall from now on call the account presented in "The Sense of Justice" and "Justice as Fairness" discussed in section 1 the *earliest* philosophy

With all the pieces of Rawls's theory in place, the just society defended is what Rawls calls a *well-ordered society*. The notion of a well-ordered society is one of the fundamental ideas of Rawls's theory. He defines it as a society in which (1) everyone accepts, and knows that everyone else accepts, the same conception of justice (2) the shared institutions of that society, which constitute its basic structure, conform to and are known to conform to that conception of justice (3) people are motivated by their shared conception of justice to maintain their just institutions and act justly towards one another.[87]

---

83 For other aspects of reasonablness, see *PL*, pp. 83, 223—225. Virtues of rightness in general are found at *TJ*, pp. 466—467/408—409, 472/413, 478—479/419. The categories of rightness and reasonableness obviously overlap in some way, but I shall not explore this matter.

84 See *TJ*, pp. 454–455/398–399.

85 *TJ*, pp. 567—577/496—505

86 See *JF*, p.196 fn17

87 See *TJ*, pp. 4–5/4–5, 453–455/397–398, *PL*, p. 35, *JF*, pp. 8–9, *CP*, pp. 233, 324

Because of their shared knowledge of the conception of justice, a well-ordered society's conception of justice is *public*, and its principles conform to the publicity condition placed on the choice in the original position via the concept of right. Public justification is available to all in the well-ordered society, which gives the relevant and objective reasons for why the society is arranged in the way that it is, rather than some other way (see further subsection 11).[88] Different conceptions of justice entail different well-ordered societies — the well-ordered society of Justice as Fairness is just one example.[89] But, by definition, in a well-ordered society, the principles of justice which organise the society are public, and not esoteric.[90]

This completes my sketch of Justice as Fairness. I have not tried to include everything. In particular, the important alterations which occur with the advent of Rawls's political liberalism are not introduced here, but rather in subsection 12.1. As I said in the introduction, however, I will include and discuss later material which is compatible with Rawls's earlier comprehensive liberalism.

Section 3: The roles of moral psychology

In outlining the structure of Rawls's overall theory, I have touched on the roles that moral psychology plays within it. There are six such roles. I do not think that they have ever all previously been separated out.[91] I view doing so as essential to any thoroughly systematic account of Rawls's moral psychology. In what follows, these six roles are introduced in turn. In the rest of the thesis, the third and fourth roles are discussed in chapters 2 and 5. The sixth role is discussed in chapter 3. The first and second roles are discussed in chapters 2, 4 and 5. The fifth role is the subject of the whole of chapter 6. Before I begin, however, I need to make some terminological distinctions surrounding the term "moral psychology".

---

88 The different levels of public justification are outlined and discussed at *PL*, pp. 66—71. That the reasons given are objective in some appropriate sense is specified at *TJ*, pp. 516—520/452—456 and *PL*, pp. 110—112, 115—116, 119—121.

89 *TJ*, pp. 454—455/398 *CP*, pp. 232—233

90 An esoteric conception of justice is one which must be kept non-public and secret in order to operate, given human psychology. Sidgwick (1907) pp. 489—490 proposed that Utilitarianism would be best served, in most circumstances, by keeping the knowledge that society is organised according to the doctrine secret. Rawls rejects esoteric morality at *TJ*, pp. 133/115, 454/398

91 Various authors indicate an awareness that Rawls's moral psychology plays multiple roles in his theory. See, for example, Krause (2008), p. 35 (though note Krause erroneously believes that Rawls has actually restricted himself to one role). Balwin (2008) p. 251 makes a similar error – see subsection 5.2 below.

3.1 Rawls's moral psychology, moral psychology, human nature, personhood

Rawls presents us, in "The Sense of Justice", chapter 8 of *A Theory of Justice*, and elsewhere,[92] with *a* moral psychology. It is the moral psychology of Justice as Fairness. It occupies a distinct section of his presentation of Justice as Fairness: the argument for stability.

But we may also want to talk of our moral sensibility, moral psychology, psychological facts, and the facts of human nature more generally. In addition, we want to know how these relate to Rawls's conception of the person. Distinguishing between these will aid our exposition generally.[93] I shall distinguish these terms in the following way.

The Person: I earlier mentioned Rawls's conception of persons or people as free, equal, rational and reasonable. Rawls describes this as a *normative* conception of the person. There are several things to be emphasised about such conceptions. First, a normative conception of the person "is to be distinguished from an account of human nature as given by natural science and social theory"[94] (see further subsection 5.2). Second, Rawls understands his conception of the person to be *a* normative conception. There are many different normative conceptions of the person. They can be "legal, political, moral, or indeed philosophical or religious, depending on the overall view to which [the conception] belongs."[95] Different societies may contain quite different conceptions of the person, different moral theories endorse or promote different conceptions,[96] and, depending on the conceptions in question, a single human being may realise several of them at once. On Rawls's view, a person, or moral person as he often says, is a human being (it is assumed) who is either capable of being free, equal, rational and reasonable, or who has realised these characteristics. I use "person" or "people" to refer to persons in the sense of Rawls's conception, and to use "human being" for persons more generically considered. This is often awkward: Rawls himself does not consistently make this distinction, and in addition, sometimes debates make things hard to phrase in these terms. But this is preferable to

92 *PL*, pp. 81—86, *JF*, pp. 195—198, *CP*, pp. 445

93 I do not claim that my stipulations of how I shall use these terms matches perfectly onto all the times that Rawls employs them. I have chosen them in order to be able to express all the distinctions I think need to be expressed in discussing his work. For times when Rawls himself mentions these broader terms, see, for example, *TJ* pp. 46/41, 137—138/119, *PL* pp. 86—88, *CP* p. 321—322

94 *PL*, p. 18 fn20. See also pp. 86—87, *JF*, p. 19, *CP*, pp. 321—322

95 *PL*, p. 18 fn20.

96 *CP*, pp. 297—299

simply using “person” rampantly.[97]

Rawls's Moral Psychology/The Moral Psychology: The specific moral psychology referred to in this subsection's opening paragraph I shall call either Rawls's moral psychology, Rawls's psychology, or the moral psychology. It includes three components. These are (1) an account of the moral character or sensibility of human beings who realise the normative ideal of the person in Justice as Fairness, (2) an account of how this sensibility relates to the rest of the person's non-moral psychology such that the moral sensibility can have sufficient control over the rest of the person's character if the person so wills, and (3) an account of processes of psychological development whereby people acquire such a moral sensibility. It is the psychology that complements the normative conception of the person just outlined. It inheres in, and is realised to some adequate level by, members of the well-ordered society. Moreover it is a psychology which, Rawls claims, human beings have the capacity to realise. In other words, Rawls hopes that human beings are able to form a well-ordered society. Perhaps even we ourselves may be such moral persons and have already realised this psychology to some extent. As I shall ultimately outline it in later chapters, this psychology will incorporate some material from outside the passages and article referred to above (see subsection 13.1). But these bodies of text will remain at its core.

Human Psychology/Human Nature: The psychological facts, and the facts about human nature, I shall use as interchangeable terms (except briefly in subsection 3.5 below). By them, I mean the broader core body of facts about human beings. This nature includes, so Rawls argues, the moral psychology, or else psychological dispositions and structures sufficiently similar to those postulated by that psychology to vindicate Justice as Fairness. Also included are many other facts about human beings more generally. Obviously, not every fact about human beings is a fact about human nature or psychology. The fact that human beings live on Planet Earth is not, for instance. Moreover, there are many different discourses and subjects which are applied to human beings. I take it there is no need for me to discuss these issues here.[98] The facts about human nature are the facts which are considered by the members of the original position, as was noted in section 2.

---

97 Rawls notes that we should also distinguish between human beings, and persons as the term is employed in the philosophy of personal identity and the philosophy of mind (*PL*, p. 31 fn34, *CP*, pp. 296—297). Any normative conception of the person is narrower than this latter notion of the person, and the class of human beings is distinct from both. Rawls postulates that any account of personal identity will under—determine what normative conception(s) of the person we should adopt, though he does not claim the two areas of debate are completely independent (*PL*, p. 31 fn34, *CP*, pp. 299—302). I do not employ this added distinction in the text, as it would make for unnecessary complexity.

98 For a brief comment on these matters, see fn68 above

Moral Psychology/Moral Sensibility/Moral Nature: Finally, moral psychology represents a subset of the facts about human nature. But it is unnecessary to identify moral psychology in general with Rawls's moral psychology, and there is no indication that Rawls believes this is required.[99] Rawls's psychology needs to be adequate for the task of defending his moral theory. But what can be properly called moral psychology in general need not be identified with his moral psychology in order for his theory to be justified.[100] People may be moral people, despite not having realised anything much like the psychology of Justice as Fairness — they can be moral in the light of a different, recognisably moral conception (see also comments under "1." in the introduction). Possession of a realised moral psychology is equivalent to possessing what I have earlier called a moral sensibility. A moral psychology, whether it is realised or not, might also be called a moral nature. It is obvious that all these terms have slightly different connotations, but I take it that I can get by without spelling them out.

Now, here are the roles.

3.2 Roles #1 and #2: Defence and explanation of psychological realisability and stability

The most prominent role (role #1) which Rawls's moral psychology plays in his theory is to argue that the principles and ideals he proposes can be *psychologically realised* by human beings in circumstances the same or sufficiently similar to ours, and are *psychologically stable*. As I said at the outset of section 1, for any theory a key question to ask is: when the normative claims of a theory telling us what human beings should be like are put forward and defended, can corresponding psychological claims also be put forward and defended, allowing us to say that human beings *can* be like that, and under what conditions?[101]

In Rawls's theory, for the most part, the moral psychology is set the more specific task of showing that when human beings have been brought up under the just institutions of the well-ordered society, they come to psychologically realise the normative conception

99 This is clear from his allowing other moral conceptions and respective psychologies. See *TJ*, p. 500/437—438, *CP*, p. 296, *PL*, p. 87

100 See, for example *TJ*, p. 578—581/506—509

101 For self—clarification here, I am indebted here to Flanagan (1996), pp. 20–22 and Flanagan, Sarkissian and Wong (2008), pp. 10–11

of the person, to some sufficient degree, and sustain this status over time.[102] I say "for the most part", as this role of moral psychology is expanded in the later politically liberal period. There it is employed in order to argue that human beings might be able to attain a well-ordered society starting from our current historical position in less just liberal democracies.[103] These two tasks can both be taken to be addressing the same, more general question: can human beings realise and sustain the well-ordered society of Justice as Fairness, under favourable conditions?

There is a lot more to be said about what this realisation consists in, and what amounts to it being sufficient. The psychological realisation of a conception of justice is obviously not simply a matter of forming the right beliefs, but also acquiring corresponding motivations. Indeed the story is even more complicated when told in full, requiring reference to sentiments, emotions, psychological developmental principles, and other more complex attitudes and traits.[104]

The phrase "favourable conditions" has been introduced. It should be briefly explained here. It may be that a well-ordered society is actually impossible for us to realise in our world. This might be for several reasons. Rawls assumes that a certain level of material well-being is required in order to be able to sustain the basic institutions of liberal democracy.[105] Our world may, conceivably, lack the resources to allow this. It should be noted that this may at best indicate that not all societies can be well-ordered. Rawls, however, thinks that the necessary material conditions are actually quite minimal, and that they most likely can be met all over the world.[106] Furthermore, however, the course of history might be such as to prevent a well-ordered society from coming about. Hostile, unjust international relations may simply make this impossible. Or it may be that the history of each individual country, and the political culture it has bequeathed, means this cannot be achieved.[107]

All three of these examples, however, rely on human beings facing *unfavourable* conditions. Because of this, they still allow that human beings, under favourable conditions, would be able to achieve a well-ordered society. Rawls is interested in the possibility that the realisability or stability of the well-ordered society could be inevitably undermined by *human nature itself*, through the sense of justice being *incompatible* with

---

102 *JF*, p. 181. See also *TJ*, pp. 144/124, 455—458/398—401, 461/404, 496—498/434—436, *PL*, pp. 140—142, *JF*, p. 88—89, 184—185, *CP*, p. 233—234, 294, 479.
103 See *PL*, pp. 86 fn34, 158—168, *JF*, pp. 192—195
104 See *TJ*, chapter 8 in general
105 *TJ*, p. 542/474—475
106 *LP*, p. 106
107 See, for example, *JF*, p. 4, *LP*, pp. 127—128

broader human nature.

For Rawls, the realisation of the moral psychology corresponding to the normative ideals and principles of Justice as Fairness must be compatible with the persistence over time of the institutions of the well-ordered society, and the persistence over time of that moral sensibility itself. The moral psychology must not only be realisable but stable. The moral psychology can fail to be so by being incompatible with the rest of human nature. The moral psychology of Justice as Fairness may be realisable, but it may happen that the other aspects of human psychology will inevitably undermine that sensibility over time. Human nature is then incompatible with the long-term realisability of the moral psychology. It should be noted that I have used human nature and incompatibility very loosely here.

Before moving on, a difference can be noted between the reasons stressed by Rawls for presenting the moral psychology in *A Theory of Justice*, and in "The Sense of Justice". In the former, it is clear that the need is to examine the prospects for the stability of the society of Justice as Fairness. But, while this concern is present in "The Sense of Justice" as well,[108] the greater focus in that paper seems to be on explaining the sense of justice which we ourselves are taken to already possess (role #2).[109] Explanation is a role of moral psychology distinct from defence. Giving an explanation of this sentiment presupposes that we already have this sentiment to some degree.[110] This idea is present in *A Theory of Justice*: that whole work also presupposes the existence of a sense of justice in persons.[111] But the concern with the defence of stability of a just society appears to have become more pressing for Rawls by the time he came to write his first book. This concern only grew of course, and contributed to the revisions of his philosophy found in *Political Liberalism* (see subsection 12.1). To summarise, moral psychology plays the roles of defending the realisability and stability of the well-ordered society, but also plays the less ambitious role of explaining our possession of a moral sensibility.

These two roles seem perfectly compatible, and do not cover the same territory. Explaining how it is that we ourselves come to be moved by a sense of justice does not amount to a defence of the viability, never mind stability, of a well-ordered society. It is this that Rawls takes to be the pressing task from *Theory* onwards.

A final comment. I have spoken here of the stability of a well-ordered society over time. But Rawls means something more specific by stability than simply the persistence of

108 *CP*, p. 104—105, 106
109 See the "second question" posed by the paper, and addressed, on *CP*, pp. 96, 99—100, and 110—112
110 *CP*, pp. 96—97
111 *TJ*, pp. 46/41

the institutions of a well-ordered society due to adequate motivations. The well-ordered society cannot be stable simply in virtue of its members being animated by any old reasons and motivations. Rather, a well-ordered society is stable because its members are moved by reasons of the right kinds — reasons of justice and reasonableness which are part of the public conception of justice governing the society. When this happens, the society is said to be stable for the right reasons.[112] This aspect of Rawls's conception of stability will be largely set aside until section 11.

3.3 Roles #3 and #4: Justification of principles, through avoiding futility and arbitration

Whether certain principles of justice are likely to yield a stable society can be conceived to be independent of the soundness of those principles. At the extreme, we can think that whether principles of justice can be realised and stable in the institutions and character of a society, and hence be matched to a psychology which meets role #1, or even whether they can be realised by human beings in any circumstances at all, is irrelevant to the correctness of the principles. It is amongst the most natural concerns in the world to want to demonstrate that human beings can live up to the standards and ideals we set forward. But if we come to believe they cannot, and if in addition we deny that "Cannot" implies "not-ought", we may judge that we should retain the standard in question.[113]

Rawls, however, argues that the realisability and stability of proposed conceptions of justice are relevant to the correctness of those conceptions. Both realisability and stability are necessary requirements for a conception of justice to be *justified*. Hence, when a moral psychology is capable of playing role #1, and defending realisability and stability, it also play the two justificatory roles: roles #3 and #4.

There are two justificatory roles because there are two aspects to any justification. First, any conceptions of justice which are impossible (or which can be expected to be impossible under any foreseeable conditions) for human beings to meet to some sufficient and society-wide degree must be discarded.[114] A moral psychology which corresponds to our theory of justice, and which is capable of being defended as realisable and stable at a

112 See *PL*, pp. xxxvii, 142—144,

113 This is the position argued for by G.A. Cohen in his *Rescuing Justice and Equality*, particularly in chapter 6. As indicated in the preface, I will do little in this thesis to arbitrate between Cohen's and Rawls's respective positions.

114 *TJ*, pp. 455/398. This also follows from the claims on pp. 159—161/137—138.

society-wide scale, can be said to allow our theory to meet the goal of *avoiding futility*.[115] Justification through futility-avoidance is role #3 that moral psychology plays. Furthermore principles of justice which, if implemented in institutions, appear to be more likely to generate motivational support from the human beings living under them should be preferred to those less likely, all other things considered.[116] Moral psychology can hence play the role of tie-breaker, and *arbitrate*. This is role #4 for moral psychology to play.

I'll make some brief comments here: first on concepts and conceptions of justice, and justification, and then on moral psychology and justification in *Theory* and in the earlier papers. The first is that Rawls does not say that conceptions of justice which fail to be realisable at all are therefore *not* conceptions of justice. He appears to hold back from this, saying potentially weaker things such as "however attractive a conception of justice might be on other grounds, it is *seriously defective* if ... it fails to engender in human beings the requisite desire to act upon it"[117] and "a strong point in favour of a conception of justice is that it generates its own support."[118] What determines that something is a conception of justice is that it can be seen to be an elaboration of our concept of justice. In section 2 I described the difference between concepts and conceptions. For Rawls, all those who understand the concept of justice recognise the need for enacting principles which ensure that "no arbitrary distinctions are made between persons in the assigning of basic rights and duties and ... the rules [of society and its institutions] determine a proper balance between competing claims to the advantages of social life."[119] Rawls does not think that a writer such as Plato did not have the concept of justice. Plato's conception of justice still fits the characterisation just given. Rather, it is simply that Plato's position is ultimately unwarranted – for its unworldliness as much as for other features.[120]

My final remark in this section, unrelated to the previous few paragraphs, is that the issue of justification does not appear to be amongst the concerns of "The Sense of Justice". There, the moral psychology is presented in order to answer certain questions about the nature of justice.[121] There is no mention of it being used in order to defend Justice as Fairness against any other view.[122]

---

115 As suggested by Rawls's language at *JF*, p. 185
116 *TJ*, pp. 456/399, 498/439
117 *TJ*, p. 455/398. My emphasis
118 *TJ*, p.177/154
119 *TJ*, p.5/5
120 *LHPP*, pp.3—4 clearly recognises the "Platonic View" of political philosophy, whilst also rejecting it. See also *TJ*, p. 454/, which rejects Plato's idea of the noble lie, which violates the publicity condition (see section 2 above).
121 *CP*, p. 96, 99—100
122 This is also noted by McClennen (1989) p. 7 fn10

3.4 Role #5: Determining the scope of justice

One role of Rawls's moral psychology present at all times in his career is the role of determining who is owed justice, and to what extent. Human beings capable of developing a sense of justice – and later also a capacity to develop a conception of the good – are owed justice, and they are owed justice equally.[123] These two capacities are in later work referred to by Rawls as the two moral powers, and they are elements of our capacity to be reasonable, and rational, respectively.[124] They are examined in more detail in subsection 4.2, and section 8. The capacity for a sense of justice, and the capacity to develop and revise a conception of the good, act as criteria for being included within the *scope* of justice.

Hence a fact about our psychology is appealed to in order to determine which individuals are owed justice.[125] This idea needs to be explained properly. What gives the criteria for determining who is owed justice is the normative conception of the person, and the drum which will be repeatedly beaten in subsection 5.2 is that normative conceptions are not the same as psychological facts (or for that matter psychological conceptions). Rather, a fact about human nature is appealed to in order to identify which individuals are owed justice given the stated criteria.

It may be wondered whether moral psychology should really be said to have a fifth role in Justice as Fairness on the basis of this. For it seems that psychology is simply indicating which particular individuals principles of justice apply to. It is, in this role, not determining anything of their content. Isn't this simply a matter of the application of the theory? But then, why should that discount this role from being a genuine role as regards the theory. Being a practical theory, we want to be able to know when and where it is applied. Given the method of reflective equilibrium, we may also be led to revise our theory when we see the practical results of it actually being applied.

Finally, as chapter 6 will investigate, the details of whether these two powers are necessary and/or sufficient to be owed justice change between different periods of Rawls's work. To begin with, the possession of or the capacity for a sense of justice is necessary and sufficient to be owed justice. By the end, this has been weakened to a sufficient criterion, and, on a certain reading still to be specified perhaps not even that.

---

123 See *CP*, p. 96, and subsequently (and with addition of the capacity to develop a conception of the good) *TJ*, p. 505/442, and then *CP*, p. 333 and *PL*, pp. 18—20

124 See *CP*, p. 312 and *PL*, pp. 18—19. See also *TJ*, p. 505/442, where the moral powers are defined, but are only referred to as "capacities".

125 *TJ*, pp. 462/404—405, 505/442, 507—508/443—444

3.5 Role #6: Constitution

Certain writers have suggested that Rawls views his moral psychology as playing what is often called a constitutive role in his theory.[126] They deny that human psychology is restricted simply to indicating the realisability and stability of a society organised according to the principles of justice, or to playing some role in the justification of the principles. They may even claim it has a wider role than determining the scope of justice. Instead, in some further sense, moral psychology, as an aspect of human psychology, is constitutive of morality as a whole.

However, what exactly is meant when someone says that moral psychology plays a constitutive role in a theory is ambiguous. In this thesis I am going to address one clear sense of "constitutes". The most obvious thing that someone may mean when they say that moral psychology plays a constitutive role in Rawls is that they mean constitutive in a metaphysical sense. The claim would be about the status of morality — that morality is solely an aspect of human psychology, and that a correct moral theory is hence a theory about a particular aspect of moral psychology. Morality is part of the world solely in virtue of there being human psychological facts in the world. This is the thesis clearly held by expressivists and certain other naturalists.[127]

Beyond this sense, it may well turn out that what is meant by saying that moral psychology plays a constitutive role can be collapsed into the previous roles we have outlined. Alternatively, perhaps there is some more subtle sense of "constitutes" which is applicable and which I have missed. This further issue will not be addressed.

The view that Rawls views moral psychology as constitutive of morality in the sense just given – that morality is nothing over and above an aspect of human psychology – is incorrect. But depending on how we interpret Rawls, the claim may be correct in a more restricted sense. Different aspects of Rawls's theory, on the face of it, seem to pull in different directions regarding this issue.[128] I believe that these apparent tensions can most likely be resolved. That resolution would produce an account of the lion's share of Rawls's metaethics. But I do not present such a resolution in this thesis. Instead, in chapter 3, I simply argue against two readings of Rawls which are mistaken on this matter. Here, however, I note the different elements of Rawls's theory which are relevant to this question.

---

126 For example, Raz (1982) pp. 186—189, Baldwin (2008), Krause (2008) pp. 28—37

127 I shall not take it to be essential for me to outline these metaethical categories here. The curious should look to Miller (2003)

128 This is argued by Fraser (2007)

Rawls may hold that moral psychology plays a constitutive role in his theory in virtue of his view, introduced in section 2, that no part of his theory should be taken to be foundational. If we assume the theory literally describes the morality of the society of Justice as Fairness in all its elements, and part of the theory is psychological theory, then part of (the) morality (of that society) will be constituted by psychological facts. Or at least, this may be one possible outcome of a thorough understanding of the coherentist element of Rawls's theory.

Alternatively, we may think, on the basis of Rawls's constructivism, that moral psychology actually plays no constitutive role in Rawls's theory. Constructivism is not discussed within the main body of this thesis, though I have attached an appendix so that the reader might know what I take to be the rudiments of Rawls's position (see Appendix I). Constructivist views see moral principles as being produced by our practical reason. The reason why this may be incompatible with moral psychology playing a constitutive role is that the account of our practical reason may not be able to be reduced to an account of an aspect of our psychology. This is not to hold that our practical reason is not, in some sense, part of human nature. The thought is rather based on the idea that accounts of practical reason, and accounts of human psychology, are distinct theories which are not reducible to each other. That Rawls endorses this kind of non-reductionism is explained in subsection 5.2.[129] We get the further conclusion that moral psychology does not play a constitutive role when we assume that, the content of morality is worked out purely by practical reason, and hence practical reason represents the foundation of morality.[130]

I believe this may represent a distorted reading of Rawls's view of the relationship of constructivism and psychology. It is true that Rawls holds that our theories of practical reasoning and our theories of moral psychology cannot be reduced down to one another. It is also true that he defends a Kantian position which holds that to attempt to found morality on "the special psychological constitution of human nature" is a form of heteronomy, whereas constructivist views are distinguished by the fact that morality is linked to our autonomy.[131] But it does not appear to be the case that moral principles are developed purely by reference to our practical reason. They are also developed by reference to the other aspects of human nature, which includes our psychology more generally.[132] The

129 Non-reductionism and reductionism regarding different sciences, concepts, properties etc. is another thing I do not feel to be my job to discuss. See Miller (2003) chapters 8 and 9 for reductive and non-reductive naturalism in metaethics.

130 Krause (2008) p. 35—36 appears to hold this interpretation of Rawls. The general view *may* be broadly correct, but most of her details certainly are not.

131 See *CP*, p. 345.

132 Observe, for instance, his response to an objection made by Schopenhauer to Kant on behalf of Rawls's constructivism: *PL*, pp. 104—107, *CP*, pp. 318—319

whole determination is thought to represent the autonomous determination of principles.[133] On closer examination, then, it may be that autonomy in a constructivist conception is perfectly compatible with psychological facts being partially constitutive of morality.

Finally, it is unclear just how Rawls's constructivism links to his coherentism — yet another piece to the puzzle which would need to be solved.[134] As I have said, I do not attempt to resolve these issues within this thesis. But these brief comments hopefully serve to show that what role, if any, moral psychology plays in the constitution of morality for Rawls is an involved question.

133 See Rawls's comments on the interdependence of conceptions of practical reason and the constructed principles. Constructing principles of justice presuppose that the conceptions, and practical reason itself, are embodied in some way (*PL*, pp. 107—108). See Appendix I for further elaboration on these tricky ideas.

134 As noted above, Fraser (2007) notes this possible tension. Relevant discussions include *PL*, pp. 95—97

Chapter 2: Moral Psychology and Justification

This chapter focuses on two of the roles of moral psychology in Rawls's philosophy linked to justification. These are the roles of arbitration, and avoiding futility. These are described in greater detail in Section 4 below, which comprises the whole of the chapter. This section also highlights an internal contradiction, or at least ambiguity, in Rawls's account of *where* moral psychological considerations enter into the account of the argument from the original position. This contradiction or ambiguity is of significance for that argument, and all similar ones.

Section 4: Justification: The Place of psychological considerations in the Original Position

4.1 Two interpretations of the place of moral psychology in justification

The two roles of moral psychology in justification are (1) demonstration that futility can be avoided for a given conception of justice, and (2) arbitration between different conceptions of justice. Moral psychology does not play both of these roles throughout the whole of the argument from the original position. That argument is split into two parts.[135] I give an initial sketch of them here. Their full complexity will be elaborated throughout the chapter.

In the first part, each party in the original position chooses between the principles of justice, on the basis of how well they protect the fundamental interest their representee has in being a free, equal, rational and reasonable person. In making this judgement, the parties take account of various psychological facts in order to assess, with regards to each proposed set of principles, whether the fundamental interests of their representee are provided for by those principles. The parties need to reach agreement on the set of principles in order to move onto the second part of the argument. In the second part, the parties as a group consider whether the society organised according to the previously chosen principles of justice would be stable. To do this they consider various psychological

135 *TJ*, pp. 144/124, 530/464, *JF*, pp. 88—89, 180—181. The fact that the argument is split into two parts is referred to in *PL*, pp. 140—141, along with a reference to I:3.6 in the same book. However, *PL* does not contain such a section, and I can find no section in the first lecture in that book which seems to obviously correspond to this topic.

facts. If the society corresponding to the principles turns out to be unstable, then they must return to the first part of the argument.

Moral psychology does not play both of its justificatory roles in both parts of the argument. In the first part, it plays both its arbitrating role, and its role in deciding whether a conception of justice avoids futility. In the second part, it does not play the arbitrating role, but only the role of avoiding futility. I shall not defend this account of the distribution of the roles of moral psychology here. Rather, at points throughout the chapter I shall indicate why it appears to be the correct reading.

I now introduce the contradiction or ambiguity in the set-up of the original position. The contradiction or ambiguity I am concerned with is over *which* psychological facts are considered in each part of the argument (I shall explain why I refer to it as a contradiction *or* ambiguity in time). There are two interpretations suggested by the text. On one interpretation, one key group of the psychological facts (which ones will be highlighted shortly) are introduced to the original position argument in the first part of that argument. On the other interpretation, this key group of psychological facts are introduced in the second part of the argument.

I will outline both interpretations briefly in this subsection. To begin the exposition of the contradiction or ambiguity, I first briefly outline the notion of rationality as employed in Justice as Fairness. I then also outline the so-called “special psychologies”. These are the common emotional dispositions and more general attitudes that are, for the most part, irrational for persons in the well-ordered society to feel or be moved by. I explain why we should start with these topics. This constitutes subsection 4.2. In subsection 4.3, I then reintroduce into the discussion the two separate roles of moral psychology in the argument from the original position, and give a short reminder of why they are present at all. In subsection 4.4, I briefly flag up the first entry of moral psychological consideration into the argument from the original position. I note that there are severe problems in interpreting what exactly is going on here. But I do not explore further. Finally, the ambiguity or contradiction in Rawls's account of the role of moral psychology in the argument from the original position is then laid out and argued for in subsection 4.5, and its significance highlighted.

To orientate the discussion, I first summarise the two different interpretations of the argument which are on offer.

Interpretation #1: The argument from the original position is split into two parts. In the

first part, the parties only consider the rational and reasonable interests of the members of the well-ordered society, who are also presumed to have an interest in possessing a sense of self-esteem or self-respect (see further sections 8 and 9). Psychology is considered in order to find out whether human beings are capable at all of being moved by such interests, and how such motivations might come about. Given this information, the parties each consider whether their representee could be expected to abide by each of the prospective conceptions of justice, and with what probability. The second part then considers motivations which are irrational when one has self-esteem, such as spite and envy. It then also considers to what extent acting justly can be viewed as rational outside the veil of ignorance and within the well-ordered society – whether acting justly and one's good are congruent in such a setting. In the first part of the argument, moral psychological considerations are appealed to in order to avoid the contract being futile *and* to arbitrate between different conceptions of justice. Specific *full moral psychologies*, each complimenting a different proposed conception of justice, are presented, i.e. different full moral psychologies are developed which complement the principles of Justice as Fairness, which complement the principle of utility, and so on. These are compared in order to try to discern which conception, assuming any are stable at all, would be more likely to generate its own support from generation to generation if its institutions were to be realised in favourable conditions. One set of principles of justice is chosen on the basis of all this. In the second part of the argument, with the moral psychology of the chosen conception of justice to hand, it is considered whether, given such a psychology and background institutions, reactive and, it is assumed, irrational motivations and attitudes will not occur to such an extent that they threaten the stability of the society over time. Arbitration is not a concern in this second part of the argument.

Interpretation #2: As above, except that the development of the various specific full moral psychologies in the light of the available facts about human psychology, and their comparison as regards stability over time, is moved to the *second* part of the argument. They are hence developed in the same part of the argument within which the special psychologies are considered. If psychological considerations are employed in the first part of the argument – it is clear that they are, in some sense (see subsection 4.4) – then they are of a much more limited nature. Again, however, as in the first interpretation, arbitration and avoiding futility are both concerns in the first half of the argument, but arbitration is not a concern in the second. The full comparison of the psychologies, then, does not play a role in arbitrating between the various conceptions of justice.

The key difference between the two interpretations is the *placement* of the development and comparison of full moral psychologies corresponding to the different sets of principles of justice. Does this development and comparison occur in the first part of the argument, or the second?

I here remind the reader what a full moral psychology consists in. This was outlined in subsection 3.1, and how such psychologies are developed will be further outlined in subsection 5.2. There it was said to consist in (1) an account of a fully developed moral sensibility, (2) an account of the development of such a moral sensibility, and (3) an account of how that moral sensibility is related to the other aspects of a person's psychology. Rawls's moral psychology is an example of such a full moral psychology.

I note one modification of use made necessary by the subject of this chapter. A significant part of (3) is the development of an account of how the fully developed moral sensibility of persons in the well-ordered society relates to the special psychologies. But where the consideration of the special psychologies enters into the argument from the original position is not what is at issue. On both interpretations, this occurs in the second part of the argument. Hence, when I refer to the full moral psychologies corresponding to different conceptions of justice, I should not be taken to be referring to the relationship between a fully developed moral sensibility and the special psychologies for this chapter. This alteration of my standard usage is only needed for this chapter. Following this chapter, I drop this alteration, and when talking of Rawls's moral psychology, or moral psychologies more generally, should be taken to be using the characterisation given in subsection 3.1.

To repeat then, Interpretation #1 holds that the development of the full moral psychologies, and their comparison, occurs in the first part of the argument from the original position. Interpretation #2 holds that this development and comparison occurs in the second part of the argument. To make the case for distinguishing between these two interpretations is a lengthy task – I ask the reader to bear with me.

4.2 Rationality and the special psychologies

The parties in the original position are described as rational choosers, who are concerned to secure the fundamental interests of those they represent – the members of the well-ordered society. The members of the well-ordered society are taken to have

fundamental, higher-order interests[136] in being free, equal, rational and reasonable persons. They have an interest in realising both the moral powers, and being able to pursue a conception of the good. Amongst their other interests is also an interest in possessing self-respect. In the well-ordered society, these interests are met, and hence the members of such a society can be said to have a certain corresponding character. This character will be laid out in sections 8 and 9. To assess how these interests can be met, the parties in the original position rely upon their ability to rationally choose.

The account of rationality used in Justice as Fairness I have not yet outlined. It will be briefly introduced here. With the account of rationality to hand, we can begin to introduce the account of the special psychologies, and then begin to assign the many aspects of the original position argument to its first and second parts.

The parties in the original position are assumed to be rational in the sense "familiar in social theory."[137] Rawls's discussion here might suggest that the sense of rationality meant here is means/end rationality, of the kind often ascribed to Hume. There is more to be said. For one, we need some kind of account of what means/end rationality is. Second, Rawls's account is actually slightly more complex than this "familiar" account. I try to get a little more precise on exactly how Rawls understands rationality in subsection 8.1. For now, we can say that though Rawls does not need to be read to accept the standard means/end account of rationality, nevertheless his account of rationality is extremely capacious. In particular, it does not place any limitations on the possible ends or interests which rational persons may have. In this respect, it is precisely like the means/end model.

One assumption Rawls incorporates into the original position which departs from the familiar idea of rationality is that the rational choosers in the original position are not subject to the special psychologies.[138] But what are these psychological attitudes? Such psychologies include inclinations towards envy,[139] jealousy, grudgingness and spite,[140] attitudes towards risk and uncertainty,[141] postures of domination and submission,[142] and so on. In saying the parties are not subject to these attitudes, what is meant that they are not moved by them, and furthermore they are not even aware of them. Initially, knowledge of the special psychologies is behind the veil of ignorance. This means the parties are not

136 *PL*, p. 73—77
137 *TJ*, p. 143/124. See Rawls's accompanying footnote for his understanding of "social theory." See also *JF*, p. 87
138 *TJ*, p. 143/124
139 *TJ*, pp. 143—144/123—124, 530—533/464—467, *JF*, pp. 87, 181
140 *TJ*, pp. 533—534/467—468, *JF*, pp. 87, 181
141 *TJ*, p. 530/464, 541/474, *JF*, pp. 87, 106—107, 181
142 *TJ*, p. 530/464, 541/474, *JF*, pp. 87, 181

aware that human beings are subject to such motivations.[143]

Now it does not seem that such motivations can always be assumed to be irrational, given how capaciously Rawls understands rationality. And indeed, for certain persons in certain unfortunate circumstances, Rawls thinks they are not.[144] What unites the special psychologies is that they can be taken to be collectively disadvantageous vices for persons who, as Rawls thinks of it, are assured of their own self-respect and self-esteem (note: this is not to say these attitudes cannot also often be irrational for those who lack self-esteem and self-respect).[145] In general, the special psychological attitudes are only good for those whose conceptions of the good, or plans of life (see subsection 8), include desires to react against another's good without any additional benefit to themselves. In *Theory*, Rawls assumes that the parties think of those they represent as "mutually disinterested," having only an interest in "their own plan of life which is sufficient for itself" such that "they have no desire to abandon any of their aims [so that] others [will] have less means to further theirs."[146] This is taken to be the hallmark of a lack of a secure sense of one's own self-respect and self-esteem.[147] Assuming a person possesses self-respect and self-esteem, if they are also afflicted by and/or act on the special psychologies, they themselves derive no benefit from this, and are most likely to be made worse off. In addition, things are most likely going to be made worse for others.[148]

There is more to be said here. It is not clear that you can make a straight inference from the fact that someone desires to injure another person's good for no additional benefit to themselves (beyond satisfying that desire) to the fact that the person lacks self-esteem or self-respect. It is also not quite clear where the interest of the members of the well-ordered

---

143 *TJ*, p. 530/464, *JF*, p. 88

144 *TJ*, p. 534/468. This may seem to be at odds with comments on p. 178/155, which state that "it is clearly rational for men to secure their self-respect" and that "self-respect is not so much a part of any rational plan of life as the sense that one's plan is worth carrying out." This second comment seems in error — it is inconsistent itself with the claim that it is rational to secure self-respect. Regarding the first comment, we have two options. One is to say that this statement is simply inconsistent with what Rawls says overall. The other is that Rawls is using the term "rational" loosely to mean in most circumstances rational. Why this is so will be made clear in subsection 8.1 This is most likely given *TJ*, pp. 400—403/351—354

145 There are two ways in which the special psychologies may be disvaluable to those who lack self-respect and self-esteem. Lacking self-respect and self-esteem does not necessarily lead to the special psychologies and reactive attitudes being part of our good. Hence to act on them may bring even those who lack self-worth no benefit. Also, those who lack self-esteem and self-respect, and who do see the special psychologies as part of their good, may yet be led into disaster by following, or even just having, their begrudging feelings.

146 *TJ*, p. 144/124—125, though the texts there talks as if the parties were the members of the well-ordered society and already have self-esteem. This isn't quite the right way to put it, in light of the later modifications of the set up of the original position which I indicated in section 2 above.

147 *TJ*, pp. 535—536/469. For Rawls's account of self-esteem, see pp440/386—387, and for the best account of self-respect, *PL*, pp. 318—319. Famously, *A Theory of Justice* does not distinguish between these two attitudes, as was observed in Thomas (1977—1978) and Sachs (1981). Recent discussions of Rawls's latter account of self-respect and self-esteem include Eyal (2005), Doppelt (2009), and Zink (2011).

148 See *TJ*, pp. 144/124—125, 532/466, 534/468

society having self-esteem and self-respect comes from. In certain circumstances, it can be rational to act on special psychologies which are meant to be evidence for a lack of self-esteem and self-respect. So having self-esteem or self-respect can't be something which is universally rational (given Rawls's account of rationality). It might be the case that the importance of self-respect and self-esteem can be established on the basis of an aspect of the freedom of persons. Part of being a free person, in the later philosophy, is said to be to conceive of oneself as a "self-authenticating [source] of valid claims."[149] But I can see possible problems with this as well. I shall simply assume that, for one who possesses a sense of self-worth, it is irrational to spite another's good.

Given this assumption about the parties, the argument for the principles of justice is then divided into two parts, as previously noted. On both the interpretations above, in the first part, the special psychologies are ignored. The parties compare the various reasons in favour of different principles of justice available to them. One set is chosen on the basis of the overall balance of those reasons.[150] In the second part, the veil of ignorance regarding the special psychologies is lifted. The parties then consider whether the chosen conception of justice will be stable, in the light of what is known about the standing human disposition to express the special psychologies.

Also considered in the second part, specifically as regards the question of stability,[151] is whether possession of a sense of justice is good for the individual in the well-ordered society. This is the question of congruence between the Right and the Good.[152] I briefly introduce this issue here. The question of congruence is whether justice is a good thing for the Just: more precisely it is the question of whether there is congruence between the perspectives of the Right and the Good in the well-ordered society.[153] Samuel Freeman sets up the problem this way: "There are two ideal perspectives in Rawls's [theory]: the original position and deliberative rationality. The former provides the foundation for judgements of justice [and right more generally]; the latter provides the basis for judgements regarding a person's good."[154] The perspective of the Right is modelled by the original position, and the perspective of the good is modelled by deliberative rationality (on this last idea, see subsection 8.1). If congruence obtains, then, from the perspective of deliberative rationality the possession of a sense of justice will be

149 *PL*, p. 32, *JF*, p. 23, *CP*, pp. 330—331
150 *TJ*, pp. 121—125/105—108, /159, *JF*, p. 95.
151 See *TJ*, pp. 567/497, *JF*, pp. 184 and 198 together, *PL*, p. 140 fn7. Of course, in what ways the well-ordered society is good for an individual is of independent interest even apart from stability.
152 See *TJ*, pp. 398—399/350, 513—514/450—451
153 See *TJ*, pp. 397—399/349—350, 513—514/450—451, 567—568/496—497
154 Freeman (2003) p. 284

recognised as a good. Rawls attempts to argue that congruence would indeed obtain in the well-ordered society.[155] He does not investigate whether congruence obtains in any other circumstances.

4.3 The two justificatory roles reintroduced

I will now reintroduce to the discussion the two roles of moral psychology in justification. This subsection will briefly recall the two ways in which moral psychology plays a justificatory role in contract theory.[156]

In contract theory, moral psychology has the role of showing how a certain conception of justice avoids futility because of the nature of the original position as a contractarian device. The choice in the original position is a collective agreement, and "for an agreement [in the original position] to be valid, the parties must be able to honour it under relevant and foreseeable circumstances. There must be rational assurance one can carry through."[157] The parties, being rationally self-interested, will simply not agree to principles when there is no prospect of stability. This follows from the finality condition (section 2). They are aware they are making the choice for the entirety of their representees' lives. When moral psychology plays this role, the concern is with certain (or near-certain) instability.

Moral psychology can also play an arbitrating role when all other considerations are tied. It may be that there are different conceptions of justice all of which are able to be stable under favourable conditions. It may be the case that the other considerations favouring them are roughly equal, or else there are good reasons on all sides and the choice is difficult to make. It may also be, however, that one of the conceptions is more likely to be stable than the others. Faced with such a choice, the parties would favour that conception.[158]

Note the relationship between the two roles. Moral psychology can first screen out

---

155 *TJ*, pp. 570—575/499—503

156 At one point Rawls comments that "in assessing conceptions of justice the persons in the original position are to assume that the one they adopt will be strictly complied with" (*TJ*, p. 145/126). This may seem to tell against moral psychology being analysed to have the roles of futility-avoidance and arbitration. Perhaps moral psychology simply has a single, more demanding role — stability guaranteeing. However, this remark simply reflects that the parties are aware they are selecting principles for ideal (strict compliance) theory (see *TJ*, pp. 8—9/7—8, and see also subsection 15.5 D). They will choose these principles knowing that human nature is such as to have a sufficiently good chance of the members of the society following them, given favourable conditions. Hence, aside from psychological considerations and the influence of fortune, they simply assume they will be strictly complied with. This is confirmed by *TJ*, p. 245/215—216. See also *JF*, p. 88—89 which explicitly says that sufficient stability is what the parties aim for.

157 *TJ*, p. 175/153. See also pp. 145/125—126, 176/153, and *JF*, p. 103

158 *TJ*, p. 498/436

those principles certain to be unstable. It can then go on to indicate, of those remaining, how likely their prospects for stability are.[159]

4.4 Initial employment of moral psychological considerations: Arguments from the strains of commitment

Let's turn to how moral psychology is first introduced in its two justificatory roles in the course of the original position argument. I begin with arguments which are principally found in §29 of *Theory*, but also elsewhere.[160] This appears to be the first place in which psychological considerations play a role in the argument from the original position. But interpreting this section is extremely complex. I simply list some of the knotty aspects of Rawls's discussion here before moving on to discuss the central ambiguity or contradiction we are concerned with.

On both of the interpretations given in subsection 4.1, most of the arguments in section §29 are in the first part of the argument from the original position. But this is not the case for all (see the next subsection). It is difficult to discern what exactly ties the arguments together. Rawls simply entitles them "Some Main Grounds for the Two Principles of Justice", and writes that they "employ the conditions of publicity and finality."[161] Some of them do not actually employ psychological considerations, but merely appeal *directly* to the fundamental interests of the representees of the parties.[162] To clarify:

---

159 Two further questions exist regarding the two justificatory roles, which I shall not be addressing here. The first is whether the two roles can definitely be distinguished. It might be thought that they can be collapsed into each other, and represented by a single judgement as to the likelihood of a particular set of principles of justice being able to be the public conception of justice of a stable well-ordered society. I accept this may be possible, but I nevertheless think these two roles, and the judgements corresponding to them, should be kept distinguished. This is because a judgement that a conception of justice is futile automatically leads to the rejection of that conception of justice, but a judgement that it is less stable than another conception does not, necessarily. Judgements about futility are judgements about absolute stability, whereas arbitrating judgements are comparative judgements of relative stability. The second issue I shall not explore in depth is whether Rawls's way of making comparative stability judgements is actually well founded. I believe in fact his arguments against utilitarianism in this respect are weak. In addition, I believe that Rawls general approach to comparing the stability if different sets of principles of justice is largely misguided. Non-futile sets of principles of justice are both quite general and quite abstract. They allow a wide range of different particular societies which could meet them. Considering the relative stability of sets of principles of justice is unlikely to come to many determinate results – at a high level of generality and abstraction, non-futile principles can be quite flexible and are capable of being specified in quite pragmatic ways (see further subsection 5.2). Any well-founded approach to making comparative stability judgements would have to make use of much more extensive empirical data. For arguments similar to those I would make on this issue, see Labukt (2009).

160 See also *JF*, p. 102—103, 124—130

161 *TJ*, p. 175/153. See also p. /155

162 See *TJ*, pp. 155—158 regarding which conception of justice best supports self-respect and self-esteem. It is argued that Justice as Fairness does, but that Utilitarianism does so less well. But some of these arguments seem to appeal to each representee's personal interest in self-respect and self-esteem. This is different from appealing to the possible consequences of a wider lack of self-esteem and self-respect, which

the use of moral psychology in the justification of a certain conception of justice relies upon *indirectly* supporting those interests, by (1) showing that the society is stable, and further (2) showing that a person could expect that they or others would be more likely to comply with that stable society's public conception of justice, with resulting benefits for securing the person's fundamental interests.

Putting aside the arguments which obviously directly appeal to supporting the fundamental interests, there are two arguments remaining in the passages I have cited. They are the strains of commitment argument,[163] and an argument for the stability-enhancing properties of self-respect and self-esteem.[164]

The strains of commitment argument raises difficulties on examination. I am unsure about is whether it is actually meant to *be* an argument which employs psychological considerations in either of the two roles. Aspects of how it is phrased suggest it may simply be an argument which appeals directly to the interests of the parties' representees.[165] Even leading this aside, there are other complications. Rather than tackle this topic here – which would require considerable space – I leave the matter for another time. Similar issues surround the arguments based on self-respect and self-esteem.

I bring up these initial arguments employing psychology (or potentially employing, for some of them) for two reasons. One is simply to indicate they are there at this place in the argument. I do not believe they are simply part of the later arguments concerning the relative stability of different societies — or if they are this is unobvious. Instead, I believe they are meant to stand on their own. Why this matter is at issue, and why it is tricky to determine, should become clearer after reading the next subsection. The second reason follows from one of the general aims of this chapter: to elaborating just how complicated the employment of psychology within the original position argument is, and how it has not been fully examined in previous work. Nor, admittedly, is it in this one. But at least the distance still to travel has been illustrated.

4.5 The ambiguity or contradiction in the place of moral psychology in the original position argument

In this section I highlight the ambiguity or contradiction in Rawls's account of how

might be expected to lead to some members of the society not being sufficiently motivated by their sense of justice. As I note below, these latter kinds of arguments may also be present, but I do not explore this further.

163 *TJ*, pp. 175—176/153—154

164 *TJ*, pp. 178—179/155—156

165 The discussion of the argument in *JF*, pp. 102—103, 128—130 in particular seems to support this.

psychology enters into and plays its roles within the argument from the original position. I first highlight three questions Rawls answers in the course of the argument from the original position. I note that Rawls is either ambiguous or contradictory about which part of the argument the first two of these three questions are answered. As noted in subsection 4.1 above, there are two interpretations of what Rawls is saying here. I provide the textual support for both of the interpretations. I respond to some possible ways one might hope to quickly resolve things in favour of one of these rival interpretations, including an interpretation put forward by Samuel Scheffler. Having satisfied myself that this is a genuine issue, I highlight the impact that both interpretations will have on the rest of the theory, and then briefly indicate the wider issues raised about the place of psychology in moral theories.

There are three different questions involved in the stability argument. The first is whether persons growing up in the well-ordered society of Justice as Fairness would acquire a sense of justice. The second is whether, on the basis of that sense of justice, Justice as Fairness would be likely to be comparatively *more* stable than a utilitarian society (or some other society) supported by that second society's own distinct sense of justice. Answering these two questions amounts to outlining the full moral psychology of Justice as Fairness and other rival moral conceptions (noting the specification I made about the use of the phrase "full moral psychology" in subsection 4.1). The third question is whether the well-ordered society of Justice as Fairness will sufficiently limit the influence of the special psychological attitudes, and be in sufficient accordance with each person's conception of the good, so as to maintain its stability. Rawls is either ambiguous or contradictory about whether the first and second of the questions given here are answered in the first or in the second part of the argument from the original position.

Here is the evidence for the second interpretation. That the full moral psychology of Justice as Fairness is worked out in the second part of the argument is suggested by its placing within Rawls's books. It is placed in chapters explicitly described as being concerned with that part of the argument.[166] In *Theory*, for instance, that the comparisons between the psychology of Justice as Fairness and other psychologies come in the second half of the argument is supported by Rawls's bald statement, at the end of the section on relative stability, that "we are in the second part of the argument."[167] This fits with another statement, made towards the beginning of his chapter, that "this argument from stability is

166 This is particularly explicit in *JF*, pp. 88—89, 103 fn26, 132, and 180. See also *TJ*, p. 504/441
167 *TJ*, p. 504/441

for the most part *in addition* to the reasons so far adduced."[168] He also writes, in the section on relative stability, seemingly referring to the comparison of stability he has just conducted, that these comparisons "are not intended as *justifying* reasons for the contract view" as "the main grounds for the principles of justice have already been presented."[169] This confirms the idea that arbitration is not a concern in the second part of the argument. If it were, then the comparisons presented could count as justifying reasons.

It may perhaps be thought these last comments refer not to the general comparison between contractarian and utilitarian psychologies, but to some brief speculations about the evolutionary origin of the sense of justice presented over the preceding two paragraphs, which are inserted into the discussion.[170] But he continues that

> At this point we are simply checking whether the conception already adopted is a feasible one and not so unstable that some other choice might be better. ... I do not contend then that justice as fairness is the most stable conception of justice. ... The conception agreed to need only be stable enough.[171]

Regarding the current exegetical question, however, there is no indication that this "checking" is limited to the comments about evolution. I believe it makes most obvious sense to see these comments as referring to the whole section, and potentially even to the whole of the chapter. I note that these comments also support the account of the distribution of the roles of moral psychology in the argument I gave at the outset to subsection 4.1. For they suggest that the comparison between Justice as Fairness and Utilitarianism made here in §76 of *Theory* is simply concerned with checking whether the chosen conception avoids futility. The aim is not to arbitrate between them.

These various passages constitute evidence for Interpretation #2. However, in the very same section, before proceeding onto the comparison, Rawls tells me that "a decision in the original position depends on a comparison: *other things equal*, the preferred conception of justice is the most stable one."[172] This suggests we are actually in the first part of the argument, and that the first interpretation is correct. And the phrase "other things equal" suggests the use of psychology for arbitration. As has been said: in the first part of the argument from the original position, the parties select a set of principles from a variety of options on the basis of an overall balance of reasons. But comparison of the

168 *TJ*, p. 455/398—399. My emphasis.
169 *TJ*, p. 504/441. My emphasis.
170 *TJ*, pp. 502—504/440—441
171 *TJ*, p.504/441
172 *TJ*, p. 498/436

different moral psychologies of these conceptions obviously presupposes that the psychologies are already worked out. So the first *and* second questions above are seemingly both to be considered in the first part of the argument. Only issues regarding the special psychologies,[173] and the congruence of good and justice, are left aside. In addition, let me repeat a line quoted just now, but now including an additional, parenthetical comment: "this argument from stability is for the most part in addition to the reasons so far adduced (*except for considerations presented in §29*)".[174] When we turn to §29, we find the arguments for the stability of Justice as Fairness, and against the stability of utilitarianism. To support this, we also get a reference to §76 — the discussion of relative stability. Hence, the arguments in the earlier passage rely upon the moral psychology of Justice as Fairness being worked out. All this constitutes evidence for Interpretation #1 — that the moral psychology is developed in the first half of the argument, and is employed in both of its roles.

One suggestion for how to resolve the issue may be to hold that, in some sense, the earlier discussion relies on a more limited selection of moral psychological considerations than the later one. We would then be able to say that the full account of moral psychology is developed in the second part of the argument, and a limited selection of psychological considerations are taken account of in the first part. This would allow us to hold interpretation #2. Rawls's general language, in the passages quoted, suggests that the later discussion adds more to the argument than the earlier one, and this may be an explanation why. However, it is very difficult to work out exactly what is to go into this thinner account, and what is not. I myself have considered the following possibilities: (1) that the earlier passages assume that the members in the well-ordered society already possess a sense of justice, and only consider the development of the sense of justice in the second

---

173 There do appear to be certain places in the argument where the exclusion of knowledge of the special psychologies appears to be forgotten. For example, at *TJ*, p. 179/156 Rawls remarks that one reason for preferring a conception of justice is that it will support our self esteem if publicly known and *thereby* help us to avoid self-contempt, *as this* "leads to contempt of others and threatens their good as much as envy does." But that self-contempt can lead to this attitude towards others is an example of a special psychology, as, by necessity, it is irrational for one who esteems themselves (note that the parties can know to avoid self—contempt, as they can view it as one of the states that can arise when one lacks self-esteem, and they know about and are moved to secure self-esteem). I do not explore this, but is suggests that Rawls's understanding of the structure of his argument is even more confused that might earlier have been thought. Another discussion which may be taken to indicate a wider inconsistency in the original position argument is that of whether excessive envy would occur in the well-ordered society of Justice as Fairness. The discussion makes reference to earlier "points in connection with stability", and his references include the discussion of *relative* stability (*TJ*, p. 536). Is Rawls here relying on his arguments against the stability of Utilitarianism? But this shouldn't be right. We are not meant to be evaluating Justice as Fairness vrs. Utilitarianism at this point. We are simply meant to be evaluating whether the conception of justice already chosen – whichever it is – is threatened by the special psychologies. I think the overall discussion suggests that Rawls is not making this mistake. He is simply making use of earlier arguments that his principles of justice support the self-esteem of the members of a well-ordered society.

174 *TJ*, p. 455/398—399. My emphasis.

part of the argument,[175] or (2) that the moral psychological considerations employed in the earlier passages are somehow "intuitive", and do not rely upon the fully worked out moral psychologies of Justice as Fairness and other conceptions.[176] But neither of these – even taken together – explain all the aspects of the text. Obviously, another possible interpretation is that the earlier argument is simply a condensed summary of the later one. But this, then, fails to resolve our difficulty. As I have shown, the later discussion appears to say conflicting things about where it plays its role in the argument. Yet another interpretation would be that psychological considerations play *no* role in the selection of principles in the first part of the argument. But then, why does the reference to §76 appear in §29 at all then? This idea seems to run against what Rawls says about the set-up of the argument in the original position in general.[177]

Another suggestion of how to resolve this interpretative problem may be this. The parties have already chosen their principles, and the corresponding moral psychology has been developed. It appears stable. But to check its stability further, we substitute for the two principles of justice some other principle(s) — the principle of utility, say. If, given our psychology and its psychological principles, the alternative conception seems to be drastically more stable, then the parties may be led to consider whether they have made the correct choice in the first part of the argument. This is supported by Rawls's comment that we are to check whether "the conception already adopted is ... not so unstable that some other choice might be better."[178] It may seem that this procedure is what is occurring over the relevant section in *Theory*. If so, our issue is resolved on the side of the second interpretation.

This suggestion is problematic in two respects. The first is that I am not even sure if it is coherent. If Rawls has already ascertained that his chosen moral conception can be paired with a suitable moral psychology, and hence the resulting well-ordered society will be "stable enough," why should seeing that pairing alternative principles with his moral psychology would be a lot more stable lead him to judge that our chosen moral conception is actually futile? He is, after all, just meant to have shown that it is not! Either Rawls is thinking about arbitrating between different conceptions of justice here – but then immediately rendering such a comparison irrelevant when he announces that a conception

175 This is suggested by elements of the strains of commitment argument (*TJ*, p. 175—177/153—154), which I earlier put aside.

176 I thought of this out of sheer desperation to make sense of the fact that Rawls tells us that the account of the development of the sense of justice and stability will provide us with reasons "in addition to the reasons so far adduced." (*TJ*, p. 455/398)

177 See, in addition to material just quoted, *TJ*, pp. 144—145/124—126, 156/135 158—161/137—139.

178 *TJ*, p. 504/441

of justice only needs to be stable enough – or else has misunderstood the relationship between ascertaining whether a conception avoids futility, and determining whether it is relatively more or less stable than other conceptions. This is more evidence that arbitration cannot play a justificatory role in the second part of the argument. In short, the statement quoted seems confused.[179]

The second point against the suggestion made above is that, Rawls does not proceed in his comparison of the respective stabilities of Justice as Fairness and Utilitarianism simply by importing the principle of utility into his developed moral psychology. Rather, he first makes the comparison by altering his moral psychological laws themselves – and hence his developed moral psychology itself —and linking those laws to alternative moral principles.[180] Certain structural features do persist in these new psychological laws. They are still based on a general psychological tendency to reciprocate (see Appendix II). But changing these laws basically amounts to putting forward a new moral psychology to accompany a different conception of justice. This is because a moral psychology always presupposes a certain set of moral principles which are being tested for their realisability and stability (Subsection 5.2 will make clear why this is so). Rawls then also contrasts a utilitarian moral psychology based upon psychological tendencies of altruism with Justice as Fairness's moral psychology, itself based on psychological tendencies of reciprocity.[181] This also amounts to putting forward a new moral psychology, linked to a different conception of justice *and* based on different psychological tendencies.

Both of these comparisons require that more than one moral psychology is being developed in order to conduct comparisons of stability between different conceptions of justice. If only one psychology was being developed, *modified*, and tested for stability, then we could say that the initial decision of the parties had already been made in the first part of the argument, and that it did not include the comparison of the relative stability of different conceptions of justice and their complementary moral psychologies. The second interpretation would then be correct. But in general, we cannot say that a moral psychology remains the same moral psychology, whilst changing the moral principles which the moral

179 In fact, the point generalises as regards using comparisons of relative stability to support the non-futility of conceptions of justice in general. Relative stability comparisons appear to be irrelevant when we are simply considering whether a conception of justice is stable to some minimal degree. If this is correct, and if we decide on the second interpretation of the argument from the original position, then this means that the discussion of relative stability is actually redundant, and Rawls can dispense with it. This may seem to be evidence for the first interpretation. But it is not, as Rawls did not recognise that he has made the mistake I have described here. Hence, even though containing an error, his text may still yield both these interpretations. Of course, we will want to resolve things one way or another ourselves on reflection. This wider matter I consider at the end of the subsection.

180 See *TJ*, pp. 499—500/437

181 *TJ*, pp. 500—501/437—438

psychology embodies. This is despite the fact that several moral psychologies can be based upon the same general psychological tendencies (see appendix II). Hence to make a comparison of relative stability, different moral psychologies need to be developed. But if different psychologies are being developed, it hence again becomes ambiguous as to whether they are being developed in the first or second part of the argument.

I have so far given the evidence for this contradiction and ambiguity in Rawls from *A Theory of Justice*. The original position argument is given again in *Justice as Fairness: A Restatement*. Are things decided one way or another there? Again, elements tell in both directions. In support of the first interpretation, arguments referring to the later moral psychology are again employed in the first part of the argument.[182] These appear similar to the arguments given in §29 of *Theory*. Against this, the section on relative stability in *Theory* is again referred to. He remarks that these are passages he would "not change substantially" and that it is "essential to see [its] role in (the second part of) the argument of the principles of justice as a whole."[183] It is unclear whether he would reform them to remove the lines I quoted above, which give rise to the second interpretation. Certainly I think this would be a substantial revision, given what I have said. In addition, Rawls at several places refers to the question of stability of the well-ordered society of Justice as Fairness being taken up in the second half of the argument, in such a manner as to leave it ambiguous as to whether the development of the moral psychology of Justice as Fairness is being postponed till then, or that it is simply the examination of the strength of the sense of justice against the special psychologies which is to be postponed.[184] This is the section of the book in which the moral psychology of Justice as Fairness is again presented.[185]

Samuel Scheffler has argued in favour of Interpretation #1, in the course of highlighting how Rawls's various appeals to stability might be seen as attempting to validate the parties' use of the maximin rule in the original position. Considering and comparing such features of the well-ordered societies of Justice as Fairness and Utilitarianism "help to show that the choice confronting the parties has features that make reliance on the maximin rule rational."[186] To do this, Scheffler claims, requires that the full psychological considerations are able to be employed in the first part of the argument.

---

182 *JF*, p. 102—103, 124—126, 127, 132

183 *JF*, p. 196 fn17. Note that remarks on pp. 186—187 regarding §76 should not be taken to be repudiating that section in *Theory*. They are simply being given as evidence that *Theory* regarded Justice as Fairness as a comprehensive rather than as a political conception of justice. Certainly Rawls's later political liberalism should not necessarily rule out relative stability comparisons, providing that it is set up in the right way so as to be confined to the development of a political conception of justice. See further subsection 14.1.

184 See *JF*, pp. 88, 103 fn26

185 See *JF*, pp. 195—198

186 Scheffler (2003) pp. 434—435.

Only by "anchoring the parties' unwillingness to accept the sacrifices associated with average utility in a carefully elaborated moral psychology and a developed account of how a workable and efficient set of social institutions could avoid such sacrifices"[187] can the use of the maximin rule by the parties seem more rational. To sustain this reading however, it must be "misleading when Rawls [states], at the end of his discussion of relative stability in [§] 76"[188] the various comments I earlier displayed in order to give evidence for the second interpretation, e.g. those on page 504/441 of *Theory*.

I can see the appeal of the reading of *Theory* given by Scheffler here. But the comments Rawls makes cannot simply be dismissed. They are not simply misleading, but are fundamentally in contradiction with the interpretation which Scheffler is trying to defend. It is not obvious that the disparity is resolved in later works. We need to resolve things in some fashion. But I shall observe shortly that there are things to be said in favour of both interpretations. In addition, however, I might note here that Scheffler's attempt to provide a better buttress for maximin reasoning in the parties might still find some support given Interpretation #2. I have already indicated that Rawls undoubtedly does, in some sense, appeal to psychological facts in the first part of the argument within the original position – though I have noted that I am unsure quite how. The argument for maximin in *Justice as Fairness*, though still ambiguous in the ways I've indicated, is still much clearer than in *Theory*, and admittedly draws on many sources.[189] Rawls also makes ready use of facts about the institutions of the well-ordered society, as found in Part Two of *Theory*. These may support the use of maximin reasoning in themselves (though I wonder whether examining how *they* fit into the whole argument might not raise similar problems to the ones we are having here).[190]

I remark on a final appeal which might be made to try to resolve the matter quickly. In subsection 5.2 I shall observe that Rawls believes that developing a moral psychology presupposes the moral principles which are being tested for their realisability. It may be thought that this decides things in favour of the second interpretation. For if the parties are meant to choose the principles provisionally in the first half of the argument, but their choice depends on a fully developed moral psychology which relies on those very principles which they are going to choose, then it seems that the argument is circular. This is not a problem, because, on the first interpretation, the parties are capable, in the first half

187 Ibid. p. 436
188 Ibid. p. 436 fn8
189 For a summary of the reasoning for maximin, see *JF*, pp. 97—100. Note also the surprising passage on p. 99 which states that "it is not essential for the parties to use the maximin rule in the original position." I do not consider here the explanation Rawls goes on to give here.
190 For example, see *TJ*, pp. 156/135, 158—159/137 *JF*, pp. 99—100, 115—119

of the argument, of developing moral psychologies for each of the conceptions of justice they are considering, and then considering which conception of justice should be chosen. The development of a moral psychology for a set of principles does not presuppose that those principles are the ones that have been chosen, though Rawls sometimes writes so as to suggest this.[191] Yet more evidence, then, that his understanding of the structure of his own theory on this point is weaker than could be desired.

Finally, before considering how this issue may be decided, I pick up on an issue left aside from the very beginning of the section up until now. I have consistently said that the aim of this section is to highlight a contradiction *or* ambiguity in Rawls's account of the placing of moral psychology within the argument from the original position. The reason I have put things this way is that, while I am fairly confident that what we are facing here is in fact an internal contradiction in Rawls's theory, I am not utterly certain. I can see various possible paths which *might* just be able to bring some kind of coherence to the whole of what Rawls is saying here. However, the task of checking such a reconstruction is beyond me at this moment. If it were to succeed, this would mean that Rawls's text is simply *ambiguous*, in a way that Rawls's actual thoughts on this matter may not have been. I admit that, from what is said here, it is a slim hope. But I leave the task up to some more committed Rawlsian than I.

This completes my account of the internal contradiction, or ambiguity (as some lucky Rawlsian workhorse may one day discover). As I have already mentioned, this issue regarding the design of the original position is of significance once raised. Comparison between the merits of different conceptions in the first part of the argument is what leads to the initial choice of the principles of justice. In the second part, as Rawls says, the goal is simply to check that the chosen principles are sufficiently stable. The basic grounds have already been presented.

What happens if we discover that a rival conception could be expected to be much more stable if we happen to compare its moral psychology with that of Justice as Fairness? It will depend on whether the development and comparison of moral psychologies is included within the first or second part of the argument. If this task is included within the first part (interpretation #1), then obviously the deliberations of the parties could be altered, and they may come to a different decision. If it is included in the second (interpretation #2), then the greater stability of that rival conception gives us no reason for the parties to revise their decision, providing that Justice as Fairness is stable enough.

Just what interpretation we side with will determine the precise importance of both

191 E.g. *TJ*, p. 504/441, *JF*, p. 88

moral psychology, and Rawls's moral psychology, within the justification of Justice as Fairness. Similarly with any other view we try to examine through the original position. On interpretation #2, Rawls's moral psychology plays a lessened role in overall justification. It merely plays the role of avoiding futility. On interpretation #1, it in addition arbitrates.

Are there any grounds to resolve this issue, or discern which of these options Rawls would have preferred? I will not investigate the matter fully here. But I will note what seem to be the major considerations. First, let's remind ourselves of the perspective from which the inhabitants of the original position are viewing the world of their representees. The parties are assumed to know only that the circumstances of justice exist (discussed further in section 7), and the general, commonly accepted facts about human nature and societies (discussed further in section 9).[192] In both cases they are motivated to find principles compatible with the ideal of persons as free, equal, rational and reasonable.

Given this aim and these perspectives, should extensive comparisons between the different full moral psychologies of alternative conceptions of justice be part of the parties' initial choice of the principles of justice? Or should only some more restricted moral psychological considerations, which perhaps assume a sense of justice in society (to try applying my ideas for interpreting Rawls from earlier), be brought to bear at this stage?

Against including the full comparisons in the first part of the argument is the idea that conceptions of justice should only be compared as regards how well they realise the ideal of a well-ordered society populated by free and equal, rational and reasonable people. If a conception of justice appears to meet this ideal adequately, given a specification made by reference to human nature (see subsections 5.2 and 9.2), and more so than others, then why extensively consider its stability comparatively? As noted, some limited moral psychological considerations could be made in the first part of the argument, and these could play an arbitrating as well as a futility-avoidance role. So to choose this option is not necessarily to reject comparing the stability of conceptions of justice altogether. The original position is not set up so as to pick the most stable conception at the cost of all other criteria. The selected conception need "only be stable enough."[193]

In opposition to this, it may be that the conception chosen, though adequately likely to be stable, is actually, to some significant degree, less likely to be stable than some other conception which was selected against in the first part of the argument. Perhaps this alternative conception of justice which was knocked out because it gave an acceptable, but overall less agreeable specification of the fundamental interests of the representees of the

192 *TJ*, pp. 126—128/109—110, 137—138/119, 200/175
193 *TJ*, p. 504/441

parties? The balance might have been shifted, however, if the full comparative stabilities of the two conceptions had been available for the parties to judge in the first part.

It should be noted that deciding this issue takes on yet more complication when we consider how we are to understand the parties employing their knowledge of human psychology. In *Theory*, Rawls stipulates that there are "no limitations" on the "general laws and theories" the parties have access to.[194] But he also notes that as "a conception of justice is to be the public basis of the terms of social cooperation" it hence "seems reasonable to say that other things equal one conception of justice is to be preferred to another when it is founded on markedly simpler general facts, and its choice does not depend on elaborate calculations in the light of a vast array of theoretically defined possibilities."[195] In later work, this theme is developed: "the general beliefs of social theory and moral psychology relied on by the parties in order to rank conceptions of justice must be ... suitably common"[196] only being those "familiar from common sense" including "the procedures and conclusions of science, when these are well established and not controversial."[197] This is all related to the publicity condition, and the idea of public reason,[198] in liberal democracy (for more see section 11).

All in all, resolution is of obvious importance. On Rawls's theory, the outcome of the original position is meant to determine the principles of justice. The decision made reverberates throughout the theory. For example, what more precise way psychological stability considerations are allowed to be taken into consideration will alter how demanding those principles are to be. The principles we derive also alter what count as ideal and what count as non-ideal situations and behaviour (see subsection 15.5 D).

This is enough to indicate that there is more to be said on these issues. Once the matter is more fully considered, we may find we do not want to side with either interpretation. Instead, we may wish to put forward a new version of the original position, or some similar device. But assuming that our new argument has a similar structure, the same issue will reappear. All that is required is that we believe that (1) principles of justice or right are to be derived from both moral presuppositions and psychological facts, (2) there are multiple possible sets of principles which meet the minimal criteria stemming from our fundamental moral ideas and psychological assumptions, and (3) given such minimal criteria, neither psychological feasibility considerations or moral considerations alone should then *obviously* determine the overall result of the derivation.

---

194 *TJ*, p. 138/119
195 *TJ*, p. 142/122—123
196 *CP*, p. 328
197 *CP*, p. 324. See also *PL*, pp. 67 and *JF*, pp. 89—90.
198 *JF*, pp. 91—92 for how public reason figures within the original position.

This is an issue regarding which committed advocates of the original position (or something like it) must decide. It will most likely concern other contractualists. Similar issues might arise for ideal observer theories. Regarding the original position, several authors have attempted to argue that stability considerations are actually the key to understanding the force of Rawls's argument from the original position,[199] or developing a more compelling one.[200] Similarly, those who would criticise Rawls's arguments based on stability need to be clear on what positions are available to Rawls or a committed Rawlsian.[201] And if one accepts that, on either interpretation, the arguments from the perspective of the original position against utilitarianism are sound, different conceptions of justice exist, waiting to step up onto the canvas.[202]

Chapter 3: Moral Psychology as Constitutive

This chapter aims to assess the arguments of two writers who claim that Rawls's moral psychology plays a wider constitutive role in his theory than any of the interpretations canvassed in subsection 3.5 suggest. Section 5 comprises the whole of the chapter. After a

199 Scheffler (2003) pp. 434—436 I have already mentioned. See also, for example, Freeman (2007a) pp. 180—188, 195—197, (2007b) pp. 90—90, Pogge (2007) pp. 117—119, 137—138, and Zink (2011)

200 For example, Okin (1989) pp. 238—249, which I view as a revisionary interpretation of the original position, against Okin herself (nothing hangs on this here). McClennen (1989) proposes dropping the original position in favour of Rawls's argument as it was originally proposed in "Justice as Fairness", which he believes can be made to work on the basis of stability considerations. The question here would be: assuming this argument works as regards the general realisability of something like Rawlsian liberal egalitarian justice, how do we then arbitrate between the various compatible liberal egalitarianisms on offer, and what role will our initial argument from stability play in them?

201 E.g. Labukt (2009)

202 For example, Richardson (2006) employs the original position in an attempt to arbitrate between Rawls's principles, and use of a primary goods metric, and Martha Nussbaum's capacities metric, and accompanying principles for a basic social minimum. The outcome of any such venture, and our assessment of it, will obviously depend on how we set up the original position regarding the matter I have raised.

brief introduction, it critically appraises the exegeses of Rawls's work by two authors – Joseph Raz and Thomas Baldwin.

Section 5: Constitution: Moral Psychology as Constitutive of Justice as Fairness

In subsection 3.5, I introduced the possibility that moral psychology plays a constitutive role in Rawls's theory. I noted that whether moral psychology does play a constitutive role is a complicated question. These and other issues relating to Rawls's meta-ethics are important. But I shall not be investigating them here. My aim is to show that the claims of two writers regarding the status of moral psychology in Rawls's philosophy are incorrect. I believe it is perfectly possible for me to achieve this aim without resolving these further issues.

I claimed earlier that moral psychology may play a partially constitutive role in Rawls's theory. But several authors have argued that either that (1) moral psychology is constitutive of the entire theory, and Rawls's theory can perhaps be completely reduced down to psychology, or (2) if moral psychology cannot be constitutive of the *entire* theory, perhaps nevertheless the theory holds that morality is founded upon psychological facts. I shall not be discussing the work of all such writers here.[203] Instead, I restrict my focus to two — Joseph Raz, and Thomas Baldwin.

In subsection 5.1, I argue that Joseph Raz's interpretation of Rawls's theory as representing "the internal constitution of our moral sense" rests on a misinterpretation of the commitments of the method of reflective equilibrium. Second, in subsection 5.2, Thomas Baldwin has argued for a decisive break between the earlier and later Rawls, such that his moral psychology plays a foundational, constitutive role in his later work. I hold that Baldwin has misunderstood admittedly difficult passages in Rawls about the relationship his theory holds to the human sciences.

5.1 Raz on our moral sensibility as morality and the reflective equilibrium methodology

Our moral sensibility is a part of human nature. It could be further held that our moral sensibility is constitutive of morality — accounting for why morality is a feature of

203 Writers I will not be discussing here include Krause (2008) and Frazer (2007)

our world. Raz, in effect, claims that this is Rawls's position. In claiming this, he misapplies the distinction I previously made in subsection 3.1 between our moral sensibility, and our moral psychology more generally. What Raz's claim should be is: Rawls's position is that our moral psychology is constitutive of morality. I will briefly outline what goes wrong in Raz's account in this regard. But my chief interest in Raz's claims doesn't come from this mistake, but from Raz's more general argument that Rawls holds that our moral sensibility is constitutive of morality. This, I hold, is a product of Raz's mistaken interpretation of the reflective equilibrium methodology. In Raz's hands, it may seem that this methodology commits us to moral psychology being constitutive of morality. But this is not the case, as I shall show.

I first take up the minor mistake about moral sensibility and moral psychology. I then follow on to the more serious mistake about the reflective equilibrium methodology.

Raz's article[204] is concerned with the interpretation and critique of Rawls's reflective equilibrium methodology. In the course of finding the most charitable and philosophically strongest interpretation, Raz finally comes to interpret Rawls as holding that a theory of morality – as developed by the reflective equilibrium methodology – is a theory of "the internal constitution of the moral sense." By internal constitution of the moral sense, Raz is clear that he means only part of what I have called moral sensibility, and certainly not what I have called moral psychology more generally.

We give an external account of our moral sensibility when we take a completely third personal stance on the attributes of that sensibility, and its development. When we give an internal account, by contrast, we are confined to an "insider's view"[205] of our moral sensibility. We are to bring to mind, through personal reflection, our own considered judgements regarding a wide variety of moral theories, and examine the results obtained by others who have done the same. We do not need to consider entirely third-personal, psychological theories of how moral sensibility develops, is sustained, or is damaged or destroyed. Nor do we need to consider third-personal accounts of the behaviour associated with this sensibility – including both behaviour a person may be aware of themselves first-personally, and behaviour they may not. Not only are theories of morality theories of the internal constitution of our moral sense, but morality *is* the internal constitution of our moral sense (I assume Raz isn't talking loosely when he says this). Morality is hence a set of judgements and beliefs, or other attitudes, as articulated and understood by the persons who have them, and does not include any psychological or biological explanations relating

204 Raz (2003a)
205 Raz (2003a) p. 187

to the formation of those beliefs or attitudes.[206]

Though Raz is right to emphasise that the final court of appeal in the reflective equilibrium methodology are our considered judgements, from our *own* perspective, on putative theories and principles,[207] Rawls would not accept Raz's divide between the internal and external theories of our moral sensibility. For one, his considered judgements do not exclude judgements on findings from psychology and the other human sciences.[208] Furthermore Raz's distinction draws an excessively crisp line between the philosophical and psychological approaches to moral theorising. I indicated in the introduction that such an understanding of the two disciplines is out of favour in many circles. Whatever we may think of that trend, Rawls would have at least some sympathy with it. He thinks it is important to say something about how, for example, "one's experiences in infancy" relate to "one's views about authority."[209] But Raz places these firmly within external theories of our moral sensibility. Most charitably, we should amend Raz's view. I will now take him to say that Rawls's theory is that morality is constituted by our moral psychology as a whole.

But Raz's view of Rawls's understanding of the status of morality, of Rawls's methodology, and the relationship between the two, is inaccurate. Raz holds that Rawls's method of reflective equilibrium is best interpreted as committing him to the idea that morality is constituted by our moral psychology. But this rests on a faulty reconstruction.

That Rawls thinks this is suggested to Raz by the way Rawls begins his exposition of his method in *Theory*:

> Let us assume that each person beyond a certain age and possessed of the requisite intellectual capacity develops a sense of justice under normal social circumstances. We acquire a skill at judging things just and unjust, and in supporting these judgements by reasons. Moreover, we ordinarily have some desire to act in accord with these pronouncements and expect a similar desire on the part of others...
>
> Now one may think of moral philosophy ... *as* the attempt to describe our moral capacity; or, in the present case, one may regard a theory of justice as

206 Ibid. pp. 186—187
207 See Scanlon (2003) p. 141—143, 147—149. To this conclusion, Scanlon's cites *TJ*, p. 46/41 ,49/, and *CP*, p. 288. But also see *JF*, pp. 31—32.
208 Daniels (1979) pp. 22—26 and (1980) pp. 48—51 are accurate on this point. See *TJ*, p. 50—51/44—45, 578—579/506—507
209 Raz (2003a) p. 187

describing our sense of justice.[210]

Raz cites the last sentence here as evidence, but quotes no further material at this point.[211] He thinks this is enough, as by this stage in the article, Raz has already presented three different interpretations of Rawls's view. He's found them all wanting philosophically. Hence, charitably, he's struck on this reading as the strongest. We need to look over what Raz has already written earlier in the article to understand why he was led to see the "internal constitution" interpretation as the only viable interpretation.

It is inessential for us to consider Raz's first two interpretations. They have obvious problems, and Raz is right to reject them. In addition, they have little connection to moral psychology in Rawls.[212] Raz's third interpretation is that the method of reflective equilibrium is to be used as a heuristic tool to sharpen our understanding of the range and structure of people's considered moral conceptions. Reflective equilibrium involves the consideration of a wide number of different moral conceptions against one's own moral judgements (section 2). Rawls suggests in "The Independence of Moral Theory" that the resolution of various debates within metaethics regarding the ontological status and epistemology of morality could be illuminated by investigating the similarities and differences between different people's judgements on moral theories and conceptions when they have reached a state of (wide)[213] reflective equilibrium. For example, Rawls holds that "it is natural to suppose that a necessary condition for objective moral truths is that there be a sufficient agreement between the moral conceptions affirmed in wide reflective equilibrium, a state reached when people's moral convictions satisfy certain conditions of rationality."[214]

Raz stresses here the distinction Rawls makes between the task of trying to reach reflective equilibrium ourselves – and hence being reflectively settled (or as reflectively settled as possible) in our normative judgements – and adopting the role of "an observer, so to speak, who seeks to set out the structure of other people's [and our own] moral conceptions and attitudes."[215] In the latter "the procedure of reflective equilibrium does not

210 *TJ*, p. 46/41. My emphasis.
211 See further Raz (2003a) p. 186
212 See Raz (2003a) pp. 181—183
213 Rawls distinguishes between wide and narrow reflective equilibrium at *TJ*, pp. 49—50/43, *CP*, p. 289, *JF*, pp. 30—31, *PL*, p. 8 fn8. Narrow reflective equilibrium is achieved when we formulate principles which are in line with our considered moral judgements. Wide reflective equilibrium requires us to bring to bear a full range of moral conceptions and theories and the arguments and justifications for them against our considered judgement. Throughout the thesis, I have assumed the wide reflective equilibrium methodology whenever talking about judgement on due reflection.
214 *CP*, p. 290. For further supporting considerations, see p. 287, and *TJ*, p. 51—52/45
215 *CP*, p. 288. See Raz (2003a) p. 184—185

assume that there is one correct moral conception [to be found]. It is [then], if you wish, a kind of psychology and does not presuppose the existence of objective moral truths."[216] These two different deployments[217] of the reflective equilibrium methodology might be called the *normative* use and the *psychological taxonomy* use.[218] They are employed for different goals. With the first use, we aim to develop our own normative outlook. With the second use, we investigate prospects for the resolution of various metaethical issues – such as the ontological status of morality, moral epistemology, the nature of moral reasoning, etc. – through an understanding of the diversity, or lack of diversity, between the conceptions people are willing to affirm on due reflection. Raz admits he doesn't really see how the latter idea might work.[219] But in any case, when the reflective equilibrium method plays this second, heuristic role, it cannot simultaneously (Raz thinks) be concerned with the truth or correctness of the various positions assembled in its psychological taxonomy.[220]

What Raz thinks is needed for this method to get anywhere is to make the connection between the taxonomy of moral conceptions and the constitution of our moral psychology. For if we assume that morality is constituted by the moral psychologies of human beings, then "knowing the structure of moral systems which survive the test of reflective equilibrium, and knowing their number and degree of similarity, may help determine whether or not morality as a whole or any part of it is a biological species—uniform phenomenon."[221] With this assumption, the methodology hence helps resolve at least one recognisable metaethical debate (objectivity vs. subjectivity) and also yields a normative theory or theories. Or at least it would do so if the assumption were sufficiently sound to yield these things. Raz thinks it is not.[222]

Raz's negative conclusion is not my concern. My objection to Raz is that his interpretation runs against the basic theme of the discussion of reflective equilibrium in "The Independence of Moral Theory" – a theme implicit, at least, in *A Theory of Justice*. This is that the reflective equilibrium methodology is a *methodology*. As the methodology it is, it aims to be non-committal between most of the different metaethical theories

---

216 *CP*, p. 289—290. Note, this is not to imply that the first use of the method presupposes the existence of objective moral truths. What determines a person's view on this matter when employing the method in its first use will simply be the judgement they themselves come to regarding the metaethical debate about the ontological status of morality.

217 Neither writer is especially clear in specifying which use they are assuming as regards their particular discussions of the reflective equilibrium methodology.

218 Scanlon (2003) calls them the "deliberative" role and "descriptive" role respectively. See pages cited in fn4 above.

219 Raz (2003a) p. 184

220 Ibid. p. 185—186

221 Ibid. p. 186

222 Ibid. p. 189—196

regarding morality. If this is true, then it cannot be that any of the particular metaethical theories is presumed by Rawls with the reflective equilibrium methodology.

Various comments from Rawls support the idea that the method of reflective equilibrium is meant to be simply a methodology. As seen, in one role it can be employed on the assumption that we are not yet in a position to resolve various metaethical debates by focusing solely on metaethics. It is often overlooked that Rawls's view in "The Independence of Moral Theory" is not that we should all stop doing metaethics, but that we should consider the possibility that progress in certain areas may be predicated on advances in normative ethics.[223] Few results in either normative theory or metaethics are ruled out, even ones in contradiction to Rawls's own considered views: "one's moral conception may turn out to be based on self-evident first principles."[224] But this could only be the case if the reflective equilibrium methodology was a methodology for approaching both ethics and metaethics.

The passages cited up to now may all be interpreted likewise. For instance, the longer, indented quotation from *A Theory of Justice* says that moral philosophy *starts* from the description of our moral sensibility. But this is perfectly compatible with arriving at either realism or anti-realism, or a subjectivist or objectivist etc. metaethic.[225] The statement that agreement in ideal wide reflective equilibrium seems a natural requirement for objectivity similarly decides nothing as regards the realism/anti-realism debate, amongst others.[226]

Raz's mistake is perhaps to have been *too* charitable, by his own lights. He indicates that he takes "the internal constitution of our moral sense" reading to be the more promising philosophically. But this is simply not Rawls's view. Rawls genuinely did think that developing comparative moral theory using the method of wide reflective equilibrium

---

223 *CP*, p. 302

224 *CP*, p. 289. Such a position contradicts *TJ*, p. 159—160/137—138, 578/506, which should be taken as Rawls's considered view, *not* presuppositions of his methodology.

225 For references to material explaining these terms, see Appendix I.

226 Scanlon's account of reflective equilibrium may also attribute false metaethical presuppositions to the reflective equilibrium methodology. He is careful to observe that the reflective equilibrium method does not presuppose a resolution to whether morality is objective or subjective: Scanlon (2003) pp. 145—146, 153. But he also states that "Rawls holds that ... considered judgements about morality and justice need not, in order to have the importance claimed for them, but the results of our causal interaction with independently existing moral properties or entities" (p. 146). He is correct that Rawls's constructivism commits him to this view (see Appendix I). But, on my reading, the method of reflective equilibrium should not presuppose constructivism or any similar kind of anti—realism, or even the kind of quietism which Scanlon himself seems to favour (See Scanlon (1998) pp. 55—56, 59—64 . For emphasising that Scanlon can be read as some kind of quietist, I am indebted to an excellent keynote talk, "Scottish Constructivism", given by Andrea Sangiovanni at the Brave New World graduate conference in 2011). Rather it may lead to any of these, but may also lead us to view our considered judgements *as* the product of causal interaction with independently existing objects. I note that Scanlon's discussion leaves it ambiguous as to whether he sees Rawls's constructivist commitment as interior or exterior to the method of reflective equilibrium.

could help to resolve long-standing metaethical issues. But it does not seem that Raz really takes this seriously. He writes that the heuristic use of the method has "little or no value in validating any moral view" seemingly because it makes so few assumptions about what is to count as validating a moral conception.[227] From a psychological taxonomy of moral conceptions by itself, he believes, we get can get nothing normative. Rawls agrees, and Raz indicates that he knows Rawls agrees.[228] But Raz does not appear to take at all seriously the possibility that the development of a wide reflective equilibrium in moral theory, even in this austerely psychological manner, may *indicate* [229] which direction to take in resolving epistemological and ontological issues, given the results in the metaethical debate so far. Or more importantly, Raz does not appear to take seriously the fact that this is what Rawls thought — for better or worse. Raz hence looked around for some further foundations or assumptions in the background which were to do the work of grounding the whole methodology. But there simply weren't any in place. Nor should there need to be any at the strictly methodological level of reflective equilibrium. Your general methodology in moral philosophy may urge you to resolve metaethical matters first, or normative matters first, or, as Rawls's does, to suspend judgement to some extent on both. But if it presupposes a solution to such topics, then it is at the very least a less—than—general methodology.[230]

Here are some final comments on this discussion. First, my arguments here address only one aspect of Raz's assessment of Rawls's methodology. I do not take myself to have addressed the many criticisms which Raz raises against that methodology, at least not directly. Nor have I commented on why Raz's own methodology, so far as I understand it, may have led him to take the approach to Rawls that he did, nor the value of this alternative methodology.[231] But in summary, Raz's account of "The Claims of Reflective Equilibrium" (the title of Raz's article) postulates at least one mis-ascribed claim. The final claim Raz identifies – that the method of reflective equilibrium reveals the structure of our moral psychology, and hence the constitution of morality – is at odds with the aims the reflective equilibrium approach must keep to in order to remain a methodology.

---

227 So I interpret Raz (2003a) pp. 185—186

228 Raz (2003a) p. 184

229 Rawls never implied anything more than this. See *CP*, p. 302

230 This is not to say that your methodology can make no presuppositions at all. Methodologies I understand to be something like your basic philosophical orientation combined with your basic philosophical tool-kit. One of the problems I have heard raised about Scanlon's interpretation of reflective equilibrium is that it is, as he admits, largely "empty as a methodological doctrine" (2003) p. 151. I agree, and I worry that Rawls's methodology (which I interpret differently from Scanlon) though somewhat more robust, is still too empty. But this is a worry for another time and place.

231 In the Raz article discussed, I think comments on pp. 181, 188, and 193, which essentially cast doubt on the possibility of morality having the shape of a moral theory such as Rawls's, are particularly characteristic of Raz. See further general themes throughout Raz (1984).

5.2 Baldwin on Rawls's two accounts of moral psychology

Thomas Balwin's recent article "Rawls and Moral Psychology" attempts to develop an account of the changing role of moral psychology throughout Rawls's work. Whilst I have argued that the psychology plays multiple roles, Baldwin's ultimate conclusion is that the later Rawls holds moral psychology to be "foundational"[232] to his theory. This is not quite to make moral psychology constitutive of Rawls's theory. But it is to make it fundamental. I contend that this claim rests on a series of subtle misreadings of Rawls's work. What Baldwin describes Rawls as describing as his moral psychology in his later work is actually just his normative conception of the person. Rawls in fact uses the term "moral psychology" in roughly the same way throughout his work. Baldwin's claim that moral psychology plays a foundational role rests on a misunderstanding of just what is and isn't that psychology. I first recount Baldwin's interpretation below. Following this, to prepare to answer Baldwin's interpretation, I highlight six ways in which conceptions of justice and the discipline of psychology interact in Rawls. I then move on to argue against Baldwin's position.

I have said that Baldwin takes the earlier Rawls and the later Rawls to mean something quite different by the phrase "moral psychology". But this is not quite precise enough. More accurately, the alteration, which is especially prominent in *Political Liberalism*, is traced back to "Kantian Constructivism in Moral Theory".[233] So the distinction is really between the majority of what I have called the early Rawls, and the end of that earlier period together with the later period.

The earlier sense of moral psychology is said to be exemplified by the account in *A Theory of Justice*. Moral psychology is "the psychology of the moral sentiments, [dealing] with an aspect of the normal development of human beings, and therefore belongs within a comprehensive account of human psychology."[234] This psychology is, admittedly, introduced to the theory "specifically in order to help with the problem of stability."[235] But Baldwin further claims that for Rawls "our psychology itself is affected by the moral value of the context in which we grow up and live"[236] such that "the development of the moral sentiments is contingent upon the moral character of [our society]." However, a "complete

232 Baldwin (2008) p. 251
233 Ibid. p249
234 Baldwin (2008) p. 249. See also p. 252. To this conclusion he cites *TJ*, pp. 489—490/428—429, and §74 in general.
235 Baldwin (2008) p. 251
236 Baldwin (2008) p. 248. He cites *TJ*, pp. 491—492/430—431

understanding of human life ... has to make room for our moral sentiments. ... Hence Rawls's early work encourages the prospect of a unified explanatory approach to human psychology which embraces both natural and moral psychology."[237] Rawls's early approach offers the prospect of seeing moral psychology as a branch of psychology in general, whilst maintaining that even if "understood only as part of the psychological theory" moral psychology must make reference to "moral notions".[238]

The later Rawls of *Political Liberalism*, by contrast, uses "the expression 'moral psychology' in a rather different way from that in which he had used it in *TJ*, as a way of capturing 'a certain political conception of the person and an ideal of citizenship'."[239] To argue for this, Baldwin cites a passage in which Rawls talks of the moral psychology being "drawn from the political conception of Justice as Fairness" rather than "originating" from "the science of human nature."[240] Baldwin holds that "largely similar accounts of the conception of the person" are found in "Kantian Constructivism in Moral Theory". So the shift is not restricted to Rawls's political liberalism. He recognises that this earlier article "does not make much use of the phrase 'moral psychology' to describe this conception of a person" but that "the phrase does occur at least once with this use."[241] In *Political Liberalism* it is "routinely"[242] described as such. Such a moral psychology is meant to capture a conception of the person which is furthermore "central to moral and political theory."[243] On this understanding, moral psychology plays a "foundational role"[244] and is hence a "philosophical moral psychology."[245] Rawls wishes to put distance between "the [philosophical] psychological assumptions inherent in his moral philosophy", and "natural psychology, the empirical science of human nature."[246] The "prospect for a unitary ... human psychology which embraces both natural and moral psychology" from the early Rawls is hence "not sustained."[247]

Against Baldwin's claim, I hold that Rawls's moral psychology is roughly the same thing throughout his career. It is a moral psychology developed to complement his conception of justice. In the earlier philosophy, though it is informed by empirical psychology in general and potentially may be included within it, the moral psychology is

237 Baldwin (2008) p. 252
238 *TJ*, p. 491/430, cited at Baldwin (2008) p. 248
239 See Baldwin (2008) p. 249, quoting *PL*, p. 87
240 *PL*, p. 86
241 Baldwin (2008) p. 249. The "use" referred to is at *CP*, p. 346
242 Baldwin (2008) p. 249
243 Ibid. p. 250
244 Ibid. p. 251
245 Baldwin (2008) p. 252. See also p. 249, which cites Rawls's slogan "Moral Psychology: Philosophical not Psychological" from *PL*, p. 86
246 Baldwin (2008) p. 252, quoting *PL*, p. 87
247 Baldwin (2008) p. 252

first and foremost a part of a moral theory. In the later philosophy, things are the same (see further section 13). The moral psychology is *never* the same thing as the conception of the person, contrary to Baldwin's reading.

I have two tasks. I need to argue against Baldwin, as he's in error, regrettably. But Baldwin's mistaken exegesis is such as to leave it open whether the position he rejects is correct. So I need to argue for this reading on its own grounds. I start with this second task. I shall illustrate that my own reading is correct for the earlier philosophy. I shall then move on to show the same for the later philosophy. Together, these two discussions will show the continuity in what Rawls means by "moral psychology." I will then show how Baldwin has misread the material.

Again, before setting out on this discussion, I attempt to bring some kind of direction to the proceedings. I here outline six respects in which moral conceptions can interact with empirical psychology to develop their respective moral psychologies. Where these six connections can be seen in Rawls work will be indicated throughout the following discussion, rather than here.

#1 Human psychology *permits but doesn't dictate* moral conceptions: As Rawls writes "human nature and its natural psychology are permissive: they limit the viable conceptions of persons and ideals, and the moral psychologies that may support them, but do not dictate the ones we must adopt."[248] Human psychology on the whole allows the moral psychologies corresponding to a number of moral conceptions to be realised.

#2 Empirical psychology helps *specify* conceptions and principles: empirical psychology plays a part in the task of specifying and developing our moral conceptions and our understanding of the associated principles.

#3 Moral psychologies can *under-specify* the relevant empirical psychology: It is permissible for a moral psychology, when included within a moral theory and hence playing the role of complementing and supporting a certain moral conception, to incorporate less detail and exactness than is required in psychological science.

#4 Moral psychologies can *optimistically interpret* empirical psychology: To some extent, where there is some doubt over empirical psychological results (or even common sense observations) with intuitively pessimistic ramifications, when developing moral

248 *PL*, p. 87. See also *CP*, p. 301

psychologies which are to complement our moral conceptions, we can interpret such data optimistically, or bracket them, so long as we can give an argument as to why they might be eliminated or mitigated in more ideal social conditions, or indicate that the pessimistic reading of the data might be a misreading or else is not conclusive.

#5 Empirical moral psychology *ultimately depends on* moral theory: After some point, progress in general empirical moral psychology depends upon progress in moral theory in general, through the laying out of the deep structure of the various moral conceptions that are recognisable in people's moral sensibilities

#6 Moral psychology is *non-reducible* to non-moral empirical psychology: Moral psychology makes use of moral concepts which cannot be reduced to non-moral ones. This is the case for moral psychology as found within moral theory, and as found within psychological theory.[249]

I now outline my general reading of Rawls's early philosophy. I shall first return to the earliest discussion of moral psychology – "The Sense of Justice." In *A Theory of Justice*, the emphasis shifts, and more detail is added.

I noted in section 1 that in "The Sense of Justice", Rawls stipulates that the psychology he puts forward is "purely hypothetical." He does not "claim that it represents what actually takes place." His aim instead was merely for it to be "reasonably plausible and to include in it only those psychological principles which are compatible with our conception of ourselves as moral beings."[250] Of the six connections that exist between empirical psychology and moral conceptions, concern with #2 and #3 is manifestly present. The moral psychology is developed in order to help address two philosophical or at least partially psychological questions: what criteria determine the scope of justice, and why are people moved to act justly?[251] In answering these questions, the moral psychology is being used to specify aspects of a prior conception of justice — the structure of the article, and its relation to the earlier "Justice as Fairness" bear this out. That the psychology is put forward only as reasonably plausible, and most likely idealised, indicates it is knowingly underspecified when contrasted to the requirements of empirical psychological science.

I've here employed the notion of *specifying* a conception. This idea was referred to

249 Note this leaves open the question of whether moral properties can be reduced to non-moral properties. I owe this point to Robert Cowan.
250 *CP*, p. 100
251 *CP*, pp. 96, 100, 110—116

in section 2. This, and the notion of *under-specification*, will perhaps be unfamiliar, so they need to be explained. When we specify an idea or conception, we take an idea which is to some extent vague, imprecise, or abstract to begin with, and look for ways in which to sharpen it and make it more precise. Rawls is obviously aware of the idea, and incorporates it into his understanding of rational deliberation.[252] This sharpening need not proceed entirely *a priori*, but can attempt to appeal to a wide range of concrete empirical examples and theory, and personal experience. It need not be assumed that specification involves the uncovering of existing but hidden sharpness, determinacy or a more precise shape of the concept or idea we are considering: this is a substantive and contestable issue regarding specification, conceptual analysis, and the relationship and distinction between the two. The topic relates to the concept/conception distinction discussed in section 2 and subsection 3.3. But it is distinct. Conceptions might be considered to be specifications of concepts. But conceptions themselves can also be further specified – this occurs, for instance, with Rawls's conception of the person (see below, and also sections 8 and 9). In general, I believe this notion fits well with the approach Rawls appears to advocate for the development and justification of moral conceptions.

To give an example of specification, suppose I am committed to promoting international justice, most particularly, for whatever reason, as regards Africa. Let's assume this is expressed by a general principle "Promote Justice for Africa." I at first apply this principle whenever I vaguely hear of anything which intuitively sounds like it might help Africans. So I give money to aid charities when I receive flyers showing starving African children, I sign petitions to have past colonial crimes recognised, I buy music by African artists, etc. I see all of this as supporting the nebulous aim "Promoting Justice for African." One day, however, I hear a African-American talking about how he believes it is demeaning that justice for Africa is always promoted by means of showing pictures of starving African children – as if the adults in Africa weren't important as well, and what's more, were simply charity cases. I realise that my previous actions may not have all actually been pursuing my stated aim. Maybe working out how to "Promote Justice for Africa" is actually quite a tricky task. I hence begin to gather more factual data, to learn African history, to try to discern which charities or campaigning organisations are actually the most effective, and are the most compatible with showing the people I wish to help proper dignity. I still subscribe to my original principle. But it is now in a much more highly specified form, and its content will have become more fine-grained than it once was.

Under-specification, by contrast, proceeds in the opposite direction. We remove

252 See *TJ*, p. 415/364—365, and also *LHMP*, p. 33

precise detail from a quite determinate conception or theory to leave only enough as to meet our needs, or else deliberately cease to incorporate further empirical information beyond a certain point. Abstraction is a form of under-specification on this understanding. These two activities can both be employed in developing the same theory. One hypothetical route Rawls may have taken to develop his theory, we might think, was to begin by abstracting from our political culture and tradition to find its fundamental normative ideas, and then attempt to specify these ideas in a theoretical and systematic way.[253]

To return to picking out the relations between empirical psychology and normative conceptions in Rawls's work, in *A Theory of Justice*, Rawls's account incorporates all six relations outlined earlier. He writes that he "want[s] the psychological account of moral learning to be true and in accordance with existing knowledge."[254] The later account in *Theory* appears to be more concerned with the strictly empirical psychological truth of the moral psychology than the earlier account. This is born out by the richer references to empirical psychology found in *Theory*.[255] Such a concern is more important, once we see the success of our overall theory as more heavily reliant on the defence of stability. Hence, this account is concerned with #1.

The added detail also plays its role in specifying the conception of justice Rawls is developing (#2). As should be clear from the accounts of justification in section 2, subsection 3.3 and section 4, developing a moral psychology which complements a conception of justice is essential. Such justification also leads to the specification of a conception of justice in the light of facts about human nature. Such a specification will encompass the fundamental moral conceptions, and illustrate how they will be psychologically embodied (see further subsection 9.2). The principles of justice and right in general are also further specified.[256]

As in "The Sense of Justice", in *A Theory of Justice* Rawls indicates that the moral psychology does not have to meet all the standards of empirical psychology to do its work:

---

253 For this understanding of specification, I am indebted to Richardson (1994) esp. chapter IV. Abstraction is dealt with at pp. 245—246. Note that I do not comment on Richardson's own illustrative account of specification of a final end in the section entitled "Rawlsian Specification of Political Ends"

254 *TJ*, p. 462/404

255 See references to theorists such as Lawrence Kohlberg, Jean Piaget, Albert Bandura, Martin L. Hoffman, A.F. Shand and others at *TJ*, pp. 458/402 fn4, 460/403 fn6, 487/426 fn19. In the earlier paper, only Piaget and Shand are referenced.

256 This general process gives out to greater complexity than I have indicated here. For example, I argue in chapter 4 that the minimal conception of the person is fixed in Rawls's theory, and stays fixed even when the conception of the person is specified, and it is shown how that conception is embodied in human beings. But this is not the case with the principles of justice. Whatever their final specification is, that represents the content of the principles, for reasons that follow from the account of stability for the right reasons in section 11. I do not enter into these complexities here.

it can under-specify (#3).

> It is impossible to take [all] the [empirical psychological] details into account; I sketch at best only the main outlines. One must keep in mind that the purpose of the following discussion is to examine the question of stability and to contrast the psychological roots of the various conceptions of justice. ... Unless the psychological account is defective in some way that would call into question the acknowledgement of the principles of justice rather than the standard of utility, say, no irreparable difficulty should ensue. I ... hope that none of the ... uses of psychological theory will prove too wide of the mark."[257]

There are multiple points within the account which indicate the tactic of optimistically interpreting human nature (#4). See, for example, comments regarding Freud's theory of moral development, both in the account of the development of the sense of justice,[258] and in the discussion of the special psychologies.[259] I leave aside further comment on this theme till section 7.

Theory also argues that moral psychology, even when considered as a part of empirical science, ultimately depends on Moral Theory, and the systematic articulation of our moral conceptions (#5).[260] This is a serious claim, not least because it is more obviously directed at empirical moral psychologists than moral philosophers. I believe that the full implications of this idea may not yet have been fully worked out, though good headway has been made by those empirical psychologists and empirically minded philosophers who are broadly sympathetic to Rawls's conception of the relationship between moral theory and empirical psychology.[261] But this topic is too far removed from the interests of this thesis. From #5 follows the weaker commitment, #6. Rawls highlights his non-reductionism later in the book, and sees the progress made in developing a substantive theory of justice as his best evidence for it.[262]

In summary, the early Rawls's moral psychology is developed primarily to complement his conception of justice. This does not prevent it from being incorporated

257 *TJ*, p. 462/404—405
258 *TJ*, pp. 489—490/428—429
259 *TJ*, pp. 539—541/472—474
260 *TJ*, pp. 491—492/430—431.
261 As pioneered by John Mikhail. See Mikhail (2011). Other work in the same paradigm includes Hauser, Marc D., Young, Liane, and Cushman, Fiery (2008)
262 *TJ*, pp. 578—579/506—507

into empirical moral psychology in general, however, and it is not framed to prevent this.

All this is roughly compatible with what Baldwin says about moral psychology in the early Rawls in ways I will not outline. I warn anyone looking over this material that Baldwin is tremendously imprecise in the way he uses the term "psychology".[263] He does not clearly distinguish between human psychology, moral psychology, moral sensibility, and the psychology of moral development as I have done.

I now move on to moral psychology in the later Rawls (including "Kantian Constructivism in Moral Theory"). The same six interactions between empirical psychology and moral theory are again admissible.

I have already quoted relevant material supporting #1 when I stated #1 initially. #6 is affirmed in the same discussion: Rawls's comment that as a "normative scheme of thought", Justice as Fairness "is not analyzable in terms of ... say, the family of psychological and biological concepts"[264] is perfectly general, and should not be read so as to be restricted to Justice as Fairness as a political conception. Rawls's presentations of his moral psychology in his later work are much briefer than that in *A Theory of Justice*, so under-specification by the lights of empirical psychology is still obviously fine (#3).[265] What I have called Rawls's optimistic approach to interpreting psychological data (#4) specifically as regards political liberalism will not be discussed in this thesis. But it is undoubtedly present.[266]

#2 is the idea that empirical psychology plays a role in specifying our moral conceptions. Rawls remarks in *Justice as Fairness: A Restatement* that he would not "change ... substantially" much of the moral psychology developed in *A Theory of Justice*.[267] The permissibility of the use of empirical psychology is complicated in the later philosophy by the strengthened requirements of public justification (section 12, subsection 13.2). But I believe it can still be employed. Hence empirical psychology can still be seen to be helping to specify moral conceptions in the later philosophy (subsection 13.1).

On the idea of #6 – the thought that much progress in empirical moral psychology depends upon our systematic understanding of the various moral conceptions – there is not much indication of Rawls himself developing this line of thought. The idea may be excluded from political liberalism, given certain facts about the public culture of the well-ordered society in question (see further subsection 13.2). But the requirements of political

263 For example, see the use over Baldwin (2008) p. 248—249
264 *PL*, pp. 87—88
265 See *PL*, pp. 81—82, 86, *JF*, pp. 195—198, *CP*, p. 445. Note the discussion in *JF* refers back to *Theory* chapter 8. See also *PL*, p. 143 fn9.
266 For suggestive passages, see *PL*, pp. lviii—lx, 86 esp. fn34, 121
267 *JF*, p. 196 fn17

liberalism do not impact on the question of whether this idea can be part of Rawls's later philosophy — understood to be wider than Justice as Fairness developed as a political conception.[268] For this claim is primarily addressed not to the reasonable citizens of a liberal democracy, but to academic psychologists. It urges them to pay more attention to moral theories, and their structural features and differences – or perhaps better, to collaborate with philosophers in doing this.

I hence hold that there is, at least, much continuity, from *A Theory of Justice* onwards, between Rawls's understanding of what ways empirical philosophy can interact with our moral conceptions, in the development of moral psychologies which are to play their various possible roles in moral theories.

We can now examine Baldwin's claims for a discontinuity in Rawls's earlier and later use of the term "moral psychology". First, Baldwin's text is ambiguous between whether he thinks that the later Rawls presents a moral psychology which is to accompany his normative conception of the person, or whether he thinks that in the later Rawls the normative conception *is* the moral psychology. The first reading is supported by Baldwin talking, in the passages quoted above, of the moral psychology "capturing" the conception of the person.[269] If this is his reading, then he and I have no disagreement, because Rawls (both early and late) *does* present a moral psychology to accompany and help specify, and so "capture" his conception of the person. However, this interpretation would not sustain the distinction Baldwin draws between the earlier, merely "stabilising" role of the moral psychology and the later "foundational" role in any way he wants.

Overall, Baldwin's article reads as if he means to identify the conception of the person with the moral psychology, or at the very least that he hasn't recognised the difference between a normative conception *being* a psychology, and a psychology *being developed to accompany* a conception. Once this distinction is made, it seems very clear that the conception and the psychology *must* be two different things. A normative conception is a body of beliefs or propositions which carry normative content. To go into the possible ontology of concepts and conceptions would be a distraction here. But whatever we say, a normative conception will undoubtedly be a different thing than a psychological conception, simply in virtue of the fact that one concerns what is normative (properties, mental states etc.) and the other concerns what is psychological. The only way to deny this would be to hold that the normative is reducible to the psychological. I have already observed that Rawls denies this is the case, and Baldwin is sympathetic with this

268 See comments by Mikhail (2011) p. 10 fn11, plus elsewhere
269 Baldwin (2008) p. 249. See also p. 250

position.[270]

How does Baldwin come to this mistaken interpretation? The problems start in his reading of §II:8 of *Political Liberalism*. His basic mistake is to fail to distinguish between the elements of Rawls's position which follow from his political liberalism, and those others which follow from his more general philosophy. Baldwin observes the title of the section – "Moral Psychology: Philosophical not Psychological" – and quotes the following passage

> This completes our sketch of the moral psychology of the person. I stress that it is a moral psychology drawn from the political conception of justice as fairness. It is not a psychology originating in the science of human nature but rather a scheme of concepts and principles for expressing a certain political conception of the person and an ideal of citizenship[271]

Baldwin reads this to mean that the moral psychology is (now) a scheme of *normative* concepts and principles relating to the conception of the person. But the moral psychology should rather be understood to be a scheme of *psychological* concepts and principles which are used to accompany ("express") a political (normative) conception of the person. As I have illustrated, Rawls believes that to be justified, any normative conception needs to be accompanied by a moral psychology which defends the realisability and stability of that conception. But this psychology is not the same as the normative scheme which it is employed to defend (see earlier comments in subsection 3.5).

What is confusing matters here is political liberalism. As will be outlined in chapter 5, any aspect of a politically liberal conception of justice needs to be drawn solely from the public culture of a well-ordered (or near well-ordered) liberal democracy. This is what prevents such a political conception's moral conception from being drawn from the science of human nature in general. Not all aspects of psychological science are admissible in the public culture (see subsection 13.2). Hence, not all aspects of psychological science can be used in developing a moral psychology to accompany the normative conceptions and principles of a political conception of justice.

Note that, putting aside such a restriction, there is still a sense in which a moral psychology designed to accompany and defend a certain conception of the person can be said to be "drawn" from that conception and not the science of human nature. But this is

270 Baldwin (2008) pp. 256—257, 261
271 *PL*, pp. 86—87, quoted at Baldwin (2008) p. 249

for the now familiar reason that such a psychology is to be part of a moral theory, and as part of a moral theory, it can be under-specified by the demands of a full psychological theory (#3). The same psychological claims which make up such a "philosophical" moral psychology can be included within an empirical moral psychology. Indeed, they can be even when those psychological claims are part of a political conception. But what cannot occur is that psychological claims which are outside the bounds of public reason are included within the moral psychology of a political conception of justice, which must be necessarily formed *within* the bounds of public reason (for these ideas, see subsections 11, 12.1 and 13.2)

Baldwin also claims that "Kantian Constructivism in Moral Theory" can be read to identify the conception of the person and the moral psychology. But the one use in the text of the term "moral psychology" can only be read in his way if you are already trying to force that reading.

> A more complex conception of the person ... together with a suitable moral psychology, is simply unnecessary.[272]

The "together" here is enough to indicate that the conception and the psychology are different things, as Baldwin seems to, inconsistently, realise. As for Rawls "routinely" describing his conception of the person *as* a moral psychology in *Political Liberalism*, I simply deny that this is the case. Even if it at times appears that he does – and I've not found any clear examples – it should now be clear why this would be mistaken on general philosophical grounds, and on the whole an uncharitable reading.

Baldwin has one direct argument for his interpretation

> An easy way to bring out the difference [between moral psychology playing a stabilising or foundational role] is to take the case of Rational Intuitionism. According to Rawls, the sparse moral psychology implicit in Rational Intuitionism is primarily one which ascribes to persons a capacity for knowledge of moral principles and a capacity for motivation by this knowledge (*PL,* p92). It is obvious that this moral psychology does little to show that it is in a person's interest to act in accordance with this motivation; but it was that task which was to be assisted by moral psychology in its

272*CP*, p. 346. The discussion the sentence is from concerns the idea that Rational Intuition finds the more complex conception of the person of Kantian Constructivism "unnecessary", but this is unimportant here.

stabilising role.[273]

My first point, easily made, is that talk of moral psychology playing a foundational role is misleading, absent evidence to the contrary. Rawls's theory, as he conceives it, does not have genuinely foundational elements. It has fundamental elements, but by this he means simply "most abstract". This was noted in section 2. Secondly, the passage trades on the idea that the role of moral psychology in Rational Intuitionism is different than in Justice as Fairness in its original formulation. But moral psychology does not have only one role in Rawls's work. From what is written above, the moral psychology of Rational Intuitionism, to use the distinctions made in section 3, will solely play roles #1 and #2, of defending, or perhaps even just explaining, the realisability of the conception of the person as motivated moral knower (as we might call them). But moral psychology has this role in Justice as Fairness as well. It is true that it has further roles, and that it is unclear whether these roles need also be present in all versions of rational intuitionism. But the plurality of roles of moral psychology in Justice as Fairness is enough to show that no simple dichotomy between moral psychology playing foundational or stabilising roles is accurate.

Baldwin's article is mistaken on a number of other exegetical points, but for the most part these, and the pseudo-problems he develops for Rawls and then solves, can be cleared up easily once this basic error is laid out.[274] I shall not take on this task myself.

273 Baldwin (2008) p. 251

274 See *TJ*, pp. 252—253, 256—257, which put forward the worry that Rawls's "new" moral psychology may leave us unable to address the problem of stability, and pp. 260—261 for Baldwin's solution to this worry.

Chapter 4: The Conception of the Moral Person and Moral Psychology

This chapter is concerned with Rawls's account of the moral person, and its relation to the psychology of the members of the well-ordered society of Justice as Fairness. Section 6 introduces the chapter, and indicates its central ambition. Section 7 then prepares for the introduction of Rawls's conception of the person by discussing the circumstances of justice — clarifying this important topic along the way. Section 8 outlines the basic features of Rawls's conception of persons as free, equal, rational and reasonable. Section 9 then presents the completed picture of persons in the well-ordered society, drawing upon sections 7 and 8, and then discusses the relationship between human psychology and this conception.

Section 6: Developing a Moral Psychology

I begin by reviewing some of the discussions which have occurred over the previous three chapters. I then summarise what I hold to be the best approach to investigating Rawls's moral psychology. This is to sketch the character of his moral person in its barest outline. All this occurs in subsection 6.1. Subsection 6.2 then introduces an initial key distinction for the way I am going to conduct the proceedings – the distinction between first- and second-order interests or ends.

6.1 Minimal ambitions

Rawls presents us with two linked moral conceptions: (1) persons as free, equal, rational and reasonable, and who are (2) living in a well-ordered society. He then develops a construction procedure which is to model the practical reasoning of such persons, in order to derive principles for how such persons would organise their society. This takes the form of a hypothetical situation of contracting, in which representatives of free and equal, rational and reasonable persons are to contract together to devise principles to protect the fundamental interests of their representees, assuming favourable conditions obtain (section 2).

In order for such a contract to be made, the parties in the original position need to be sufficiently assured that the terms of the contract will be abided by. Hence, they attend to the facts of human psychology, in order to see which principles of justice will be

realisable and sufficiently stable at a society-wide scale. The conception of justice which best meets the representees fundamental interests, in part through having the best chance of being stable, is the conception judged to be justified overall (section 4).

However, whilst developing a moral psychology which corresponds to certain moral principles, the very same moral psychology can play a role in specifying the moral principles and developing their content. The moral psychology developed also presupposes the moral conception of persons as free, equal, rational and reasonable, and specifies how this is to be embodied in the members of the well-ordered society (subsection 5.2).

This chapter is concerned to outline the central aspects of the conception of the person in the well-ordered society. I remind the reader here of the restrictions of my discussion already put down. I have forsworn the use of extensive empirical data, and debates whose resolution relies on such data will be unable to be decisively settled. This might be thought to be a weakness of my discussion. However, I believe it to be a necessary precursor to the judicious use of empirical data to sketch a theorist's fundamental normative ideas in their most *minimal* details.

Outlining Rawls's ideal of personhood requires us to be as careful to indicate what is *excluded* by that ideal as what is *included*. I want to indicate which aspects of the account of the persons in the well-ordered society are elements which follow from Rawls's underlying normative conception of the person, and which parts follow from his assumptions about human nature. Given Rawls's assumptions, the normative conception of the person is something which must be realisable and compatible with human nature (subsections 3.2, 3.3). But the relationship between normative and psychological claims is different in these two aspects of his theory. The normative conception of the person is a non-revisable standard which human beings need match up to (as is the conception of the well-ordered society, I feel). The rest of the normative content of the theory – the principles of justice, the account of institutions, and the full moral psychology of the members of the well-ordered society of Justice as Fairness – is specified with reference to human nature, and hence can be revised in the light of human nature.

Once we have a clear picture of what is implied purely from the conception of the person (together with the conception of the well-ordered society — see section 2 and section 11), then we can get a clearer picture of what in the rest of the theory is based on Rawls's assumptions about human nature. My aim is not to assess Rawls's assumptions about human nature here, as noted in the introduction to the thesis. To stress why: the actual content of Rawls's moral psychology is both complex and expansive, and forms a tightly unified system. Critically examining it, particularly in the light of empirical

evidence, would be a task which would take up a whole other thesis, and would be the task of some more empirically-orientated study. My aim is rather to set the groundwork for the correct philosophical orientation towards this ambition.

The ambition of sketching the character and development of the persons in the well-ordered society in their most minimal details rests on being careful to distinguish between different categories of interests, desires etc. possessed by the members of the well-ordered society. The next section introduces the idea of the *first-order* and *second-order* interests, aims, ends and desires of the members of the well-ordered society. Two further categorisations of interests – intrinsic vs. instrumental, and non-public vs. public – are also introduced, to be elaborated over the coming sections. Eventually, these categorisations of interest will be combined to help give us the minimal account of the conception of the person.

6.3 First- and second-order interests

The distinction between first- and second-order interests, aims, ends, preferences, desires, concerns etc.[275] is a common one.[276] I here simply reproduce Rawls's own discussion of altruism. It introduces the idea clearly enough.

> There is ... a peculiar feature of perfect altruism that deserves mention. A perfect altruist can fulfil his desire only if someone else has independent, or first-order, desires. To illustrate this fact, suppose that in deciding what to do all vote to do what everyone else wants to do. Obviously nothing gets settled; in fact, there is nothing to decide.[277]

What holds of altruism, we shall find, holds of most of the other various basic capacities and powers contained in the conception of the person. The interests which are served by our employment of these capacities are chiefly our second-order (or third-order etc.) interests — interests about our other interests. Our job will be to discover the first-order interests which allow these various higher-order interests to become active. Unless we can

275 Though all these concepts (ends, desires etc.) are subtly different, I take that my reader is familiar with the general kind of things they all describe, and that there is no need for me to discuss them. I use this breadth of terms in order to convey that I am not talking about mental states with specific phenomenologies, as might be suggested (to some) if I just used "desire", as Rawls usually does. Note there are also issues regarding how aims and ends relate to motivation which I am glossing over.

276 The classic statement of the idea is still Frankfurt (1971)

277 *TJ*, p. 189/165

ascribe suitable first-order interests to the members of the well-ordered society, we cannot say *anything* about what they will actually do. It is only in the presence of first-order ends that second- or higher-order ends can be pursued.

Three short notes of clarification here. First, there are two different kinds of second- or higher-order interest. There are those such as the one I have just mentioned: altruism. Others include reasonableness and rationality. What bands these interests together is that they are interests in other interests being ordered in a certain way, rather than being interests in other interests having certain determinate contents. An altruistic interest is an interest in others' interests being met, whatever those interests are. A second kind of second-order interest does constitute an interest in a person acquiring a further more concrete interest. An example might be having an interest in having an interest in becoming vegetarian. I want to want to be a vegetarian. However, I also want a bacon sandwich. If my second-order interest was realised, the resulting first-order interest would be in conflict with my existing first-order interest in quite an obvious way.[278] In this chapter, I am uninterested in this latter category of interests, desires etc. Rawls conception of the person only includes higher-order desires of the former category. Hence, I ask the reader to put the latter type of higher-order desire or interest out their mind, and understand my use of higher-order interest to only refer to the former.

Second, simply because a certain power or capacity serves a second-order interest does not prevent the employment of that same capacity being the subject of a first-order interest. Indeed, Rawls relies on this idea in order to fully explain the value of the well-ordered society, as we shall see. To illustrate the general idea here: altruism moves me to help the interests of other people to be met. The aim of altruism is to help others. However, I may also want to be an altruistic person. Of course, if I am some Robinson Crusoe, this particular interest of mine will not be met. Alternatively, though I am an altruistic person, and (we'll assume) have an interest in meeting the interests of others, I may not have an interest in having my interest in meeting the interests of others. I may wish not to have an interest in altruism. What holds for altruism, and other's interests, holds for other capacities which are orientated towards our own interests. We need to find the first-order interests of the members of the well-ordered society which are *not* interests in using their second-order capacities.

This last comment leads on to a more general point. I aim to distinguish between the higher- and first-order interests, desires etc. of the members of the well-ordered society.

[278] I am unsure as to whether these is a sharp a distinction between these two categories of higher-order interest as I have indicated. But I ignore this complication – the dichotomy I assume here is enough for my requirements.

But this will not be enough to give the full account of the conception of the person met by such individuals. As mentioned in the previous subsection, there are other categories of interest which must be ascribed to the members of the society in order for our account to be complete. We are hence not simply looking for the first-order desires of the members of the well-ordered society, but a specific class of first-order desires which falls into other categories as well. What these other categories are will be revealed in due course. The class of first-order desires, interests or ends we are seeking we will simply call the *key* desires, interests or ends. They are the key first-order desires or interests as, until we ascribe them to the members of the well-ordered society, we cannot say that the members of the society will actually be motivated by anything.

Third and finally, here we might ask whether the capacities, powers and characteristics we ascribe to the members of the well-ordered society will tell us anything before we hit upon the key first-order desires. What we shall find is that they serve to *restrict* what the first-order interests of the members of the well-ordered society could be, without directly indicating what their actual first-order interests will be. This overall account, as I develop it through the chapter, may seem bound to be too indeterminate. I do not believe this is the case. Our minimal conception of the persons in the well-ordered society will be found to contain the appropriate first-order interests and desires. But they will still leave the conception of the person quite minimal.

We are hence to look for a subset of the first-order desires of the members of the well-ordered society. I say subset, as the desires we are looking for must also fall under two further categories. They must be non-public, as opposed to public, and they must be intrinsic, as opposed to instrumental. How exactly these further categorisations are to be understood will be introduced in the course of the discussion in section 7.

Section 7: The Circumstances of Justice

This chapter presents an account of Rawls's conception of the person, as this conception would be psychologically realised in the well-ordered society. The well-ordered society, however, exists under the circumstances of justice. The account of these circumstances is part of what is needed order to give the minimal account of what the members of the well-ordered society will be like.

I first recount Rawls's account of the circumstances of justice. I then comment on whether his account of the circumstances of justice needs to be modified. I also try to get clear on what the relationship between the circumstances, social cooperation, and justice

exactly is. I argue all three of these should be sharply distinguished. Clarity on this matter will allow us to criticise the oft-repeated idea that the circumstances represent problems which justice remedies. This leads on to a brief, introductory discussion of one of Rawls's major themes — that of attempting to outline that the need for liberal democracy, in the face of diversity, is not solely a regrettable fact.

Rawls announces that the circumstances of justice are conditions under which social cooperation is both possible and necessary. Social cooperation is possible, as all have interests which can be realised through social cooperation. Social cooperation is also necessary to meet those interests. It is the role of justice to distribute the goods of social cooperation – and hence meet such interests – fairly.[279]

In his characterisation of the circumstances of justice, Rawls departs primarily from Hume.[280] In the initial presentation of the circumstances of justice, the characteristics of these circumstances are as follows. Humans exist within a shared geographical territory. There is a rough equality in their physical and mental powers, such that no single individual, or coalition, is invulnerable to having their plans thwarted by the rest. Natural and other resources are moderately scarce. Humans do not live in a cornucopia, such that all needs, desires and interests can be satisfied. Nor do they live in a world so barren that cooperative schemes must break down. Together, Rawls calls these characteristics the objective circumstances of justice.

Furthermore, the human beings within such circumstances have interests which, while partially overlapping, are also to some extent in conflict. Each has their own conception of the good, and the demands of all the conceptions taken together cannot be fully met under the moderate scarcity, and divergence in conceptions of the good, that is faced. No single conception of the good is shared by all. In addition each individual possesses interests which are not interests in others' interests, i.e. are not interests in helping or hindering another's good. Finally, these individual's knowledge is incomplete, and their use of their intellect falls short of perfect. There is hence disagreement: scientific, philosophical and religious. The latter conditions listed here are the subjective circumstances of justice.[281]

Rawls's later work appears to alter, and in certain ways weaken, the characterisation of the circumstances. In “Kantian Constructivism in Moral Theory”, he

279 *TJ*, pp. 4/4, 126/109

280 See *TJ*, p. 126/109 fn 3. For the first introduction of these circumstances into Rawls's work, see *CP*, pp. 52—53. The first use of the name “circumstances of justice” is found at p. 178, in “The Justification of Civil Disobedience”. For Hume's original discussion in the *Treatise*, see bk. 3, part 2, sec. 2, paras 5—7, 16. For the discussion in the *Enquiry*, see sec 3.1

281 See *TJ*, pp. 126—127/109—110 for both the objective and subjective circumstances.

allows the possibility that the moderate scarcity of natural resources with respect to our needs may perhaps one day be overcome. This, however, is not presumed to remove or eliminate the conflict between conceptions of the good, nor remedy the limits to our knowledge and reasoning.[282] The concession is not repeated in *Political Liberalism* or *Justice as Fairness: A Restatement*, but it is perfectly compatible with their content. In these works and elsewhere, Rawls stresses that liberal democracy, by its very nature, is marked by a pervasive pluralism of reasonable outlooks, doctrines and conceptions.[283] Other remarks strongly suggest that human nature is marked by pluralism in general – a pluralism which can be suppressed only by illegitimate force and coercion, and never eliminated.[284] It is this plurality of conceptions of the good, and divergence in views, arising from what are called the burdens of judgement,[285] which is stressed above all. This idea will be returned to in subsection 12.1, when we come to discuss the distinctive ideas of Rawls's later period. The rough equality of human beings is not specified in any of the characterisations of the circumstances of justice after *A Theory of Justice*.[286] Some writers have pointed out that this simply does not obtain in our world. Some individuals – whole societies at times – have been in a completely vulnerable position compared to their aggressors.[287] It is likely that no one, not even the most cautious and well-established dictator, or hermit, has ever been entirely invulnerable to others' aggression. In summary, moderate scarcity, and human beings' mutual vulnerability, are de-emphasised and perhaps even rendered inessential as Rawls's thought progresses, whilst the plurality of outlooks is placed to the fore.

What relation does justice bear to the circumstances of justice? Rawls gives us two answers. In *Theory*, the circumstances of justice, when they give rise to social cooperation, indirectly give rise to the need for justice, and similarly indirectly give the role that justice must play within a cooperative scheme.[288] Social cooperation makes for mutual benefit to all, but conflicts of interest persist. Principles of justice are needed in order to arbitrate the various conflicts of interest, and to distribute cooperative benefits. Outside the circumstances of justice, then, there would be "no occasion for the virtue of justice, just as in the absence of threats of injury to life and limb there would be no occasion for physical courage."[289]

---

282 See *CP*, p. 329
283 See *PL*, pp. xvi—xvii, xxiv—xxv, 36—38, and *JF*, pp. 3—5, 33—35
284 *PL*, p. 37, *JF*, p. 34
285 On the burdens of judgement, see *PL*, pp. 54—58, *JF*, pp. 35—36
286 For noting this I'm indebted to Stark (2009) pp. 79—81
287 See Barry (1995a) pp. 40—41, Stark (2009) pp. 83—84
288 *TJ*, pp. 4/4, 126/109
289 *TJ*, p. 128/110

In the later specification in *Political Liberalism*, the circumstances of justice are assumed in order to render the idea of the well-ordered society "suitably realistic."[290] In *Justice as Fairness: A Restatement* – another account from the later period – they are said to reflect "the historical conditions under which modern democratic societies exist" (and, we might add, will be expected to exist in perpetuity).[291] I believe that the earlier and later accounts can be related to each other in the following way. In the later philosophy, Rawls is concerned only with developing a conception of justice suitable for a modern liberal democratic society. Without these assumptions about the objective and subjective circumstances of justice, we may end up developing a conception of justice ostensibly for such a liberal society, but which in actual fact is unrealisable. His later philosophy requires that he retreats from the earlier claim that it is only under such circumstances that justice will be called for, for reasons which follow from the limitations about what conceptions of justice are allowed to claim in the later philosophy (see subsection 12.3). Granted this, however, we can still say that the circumstances of justice bear the same relation to justice as in the earlier philosophy.

The idea of the circumstances of justice, as employed by Hume, Rawls, or anyone else, has been widely discussed. I briefly comment on some aspects of the debates here. First, discussions by certain writers broadly sympathetic to Rawls's employment of the circumstances of justice suggest they should be revised or added to in certain ways. Take, for example, Peter Vanderschraaf's argument that the account of the circumstances of justice Rawls accepts does not render social cooperation and justice possible, but rather, upon examination, is formally equivalent to a Hobbesian state of nature. If so, justice will be *impossible* without further conditions obtaining.[292]

Two points can be made here: first, additions to the circumstances of justice needed to make social cooperation possible are unproblematic for Rawls, providing that something like original aspects of the circumstances are retained. As Rawls himself adapted Hume's account, and then seemingly further tweaked his own account, debates about further alterations do not necessarily pose a great danger to him.[293] So long as the conditions we end up with for social cooperation to be possible and necessary do not radically depart from or transcend Rawls's account, then his account of the circumstances can be retained, which is the important thing from his perspective.

Second, Rawls himself does not think that the circumstances of justice are

290 *PL*, p. 66
291 *JF*, p. 84
292 Vanderschraaf (2006) pp. 321—329
293 Alterations of his account may include adaptations as well as additions. For example, Ci (2006), p. 45—60, argues that the subjective circumstances should be re—conceived

sufficient for justice. They only render it possible and necessary given that humans can be motivated by a sense of justice. The circumstances of justice are not sufficient for social cooperation. To see this, the work "possible" and "necessary" are doing in the specification needs to be further clarified. In saying that the circumstances of justice render social cooperation necessary, Rawls should not be taken to mean that these circumstances *necessitate* social cooperation. This is to ascribe an explanatory role to the circumstances of justice, as with the account in Hume's *Treatise*. Such an explanation, however, requires additional assumptions about the extent to which human beings are rational and reasonable, and are able to develop publicly recognised rules and procedures. Rawls considers these further matters elsewhere.[294] As he states them, the circumstances of justice do not amount to an explanation of why social cooperation occurs, but are rather simply conditions which make it possible. The circumstances of justice then render social cooperation necessary in a prudential sense. The majority, at least, of the interests and needs of human beings require social cooperation to be fulfilled. There is no other way.

Often, the circumstances of justice are described as giving us the role of justice. This is not quite right. More accurately, the role of justice stems from the features of social cooperation which social cooperation inherits from the circumstances of justice. Rawls writes

> principles are needed for choosing among the various social arrangements which determine [the] division of advantages and for underwriting an agreement on the proper distributive shares. *These* requirements define the role of justice. The *background conditions* which give rise to these necessities are the circumstances of justice.[295]

It seems clear here that the need to arrange our cooperative scheme is what gives rise to the need for justice. Of course, that there can be a cooperative scheme, and the particular issues which must be settled regarding it, presupposes the existence of circumstances of justice. But the circumstances of justice only give rise to justice indirectly. They do constrain its role. But they do not, as has often been argued, determine its content and scope.[296]

The reason why this distinction is important is because it allows the correct

294 *TJ*, pp. 142—145/123—126. On the need for public rules for cooperation, see *PL*, p. 16
295 *TJ*, p. 126/109. My italics.
296 For example, Hubin (1979) pp. 9—10, 21—24, Nussbaum (2006) pp. 103—104, 119, Barry (1989) chapter V, esp. pp. 179—189

perspective on the circumstances of justice to be adopted. A characterisation often proposed is that the circumstances of justice represent the *problems* that justice is to remedy.[297] But this is, at most, only half the story. It is at odds with various aspects of the classic historical accounts. Hume refuses to engage with the debate as to the virtue or viciousness of the aspects of our persons which give rise to the need for justice on his theory.[298] Of the characteristics of humankind which give rise both to the State of War and the State of Peace, Hobbes, with characteristic melody, writes "the Desires, and other Passions of man, are in themselves no Sin."[299] The presuppositions of social cooperation, then – the interests and means for meeting those interests – do not represent problems in themselves. Rather they give rise to problems in social cooperation. But at least some of these problems only arise in the absence of justice. The better rendering is this: the circumstances of justice *can* give rise to problems in social cooperation. But at least some (note: not all) aspects of the circumstances which might otherwise lead to problems do not lead to problems at all *if* justice is attained. Circumstances which give rise to problems in unjust societies can in fact give rise to great benefits in just societies. For example, religious diversity in the past led to civil war. But now, some religions at least appreciate the greater diversity which liberal democracy allows to be publicly expressed: for instance, believing that dialogue with those of other faiths allows them to understand their own faith in a deeper way.

What is going on here needs to be more precisely outlined. First, let's distinguish between interests which are instrumentally served by justice, and interests in justice, and political society more generally, which see justice as intrinsically valuable — valuable for its own sake. This is a further distinction between categories of interests, desires etc., in addition to the distinction between first- and second-order desires, interests etc. given above in subsection 6.2. The "problems" I have just spoken of should be thought of as failures, or potential failures,[300] to meet both these types of interest – intrinsic and

297 See Ci (2006), p. 45, Hope (2010). By contrast Vanderschraaf (2006), pp. 321, 332—333 observes that the circumstances of justice give rise to problems, rather than being problems. But even this isn't quite right. The circumstances of justice do contain elements which are unavoidably problems. But, with regards to those aspects of the circumstances which are not unavoidably problematic, if we go straight from the circumstances of justice to a just and beneficial social arrangement, on my analysis (see this paragraph and the next) it is odd to say that justice is remedying a problem for us. We should say: we would have had a problem, to which justice would have been the solution, but as we got justice straight away, there never was a problem.

298 See Bk 3, part 2, sec. 2, para 13 of the *Treatise*

299 *Leviathan*, Part 1, chpt. 13, para 10. It is often remarked that Hobbes's account of the natural passions and equality of human beings (though not, necessarily, their "naturall condition", the state of nature – see the reference to Vanderschraaf above) is the ancestor of Hume's account of the circumstances of justice.

300 The word "problems" as used so far has been ambiguous between actual existing problems, or problems which have been remedied by justice.

instrumental. In all societies – just or unjust – many interests always remain unsatisfied.[301] So there are always problems in life. An unjust society does not, of course, pose a "problem" to those who benefit from it, and have no wish to see things change. Rather, such social arrangements are a problem for others, particularly for those who see value in justice. The just society, then, meets many of the instrumental interests of all, in a fair way, and also meets the ends of those who thirst for justice for its own sake.

The claim that a just, well-ordered society is intrinsically valuable is one of the key claims which Rawls wants to argue for. Rawls is at pains to argue that the just liberal society is not equivalent to a "private society", whose members are not assumed to have any shared ends realised by their political institutions.[302] Following his later philosophy, I think it is acceptable here for us to call the shared, intrinsically valuable political interests or goods met by the well-ordered society public ends, and the "private" ends – instrumentally met by the well-ordered society – non-public ends.[303] It is part of Rawls's definition of a well-ordered society that it is valuable in itself.[304] This is in addition to saying that it is valuable instrumentally in that it allows everyone to realise their good to some adequate and fair extent. Note that, in holding this, Rawls does not need to be taken to be saying that there would be no worth in a society which did *not* live under the circumstances of justice, and hence which did not have to realise justice. This thought is unnecessary. Rawls is not saying that we should bring about justice even if we do not need it. All he is saying is: here is justice, and it has intrinsic worth.

This idea leads on to an important aspect of Rawls's theory and philosophy in general. (We have already briefly touched on an aspect of this in subsection 5.2.) Rawls appeals to us to try to see our inescapable historically grounded human condition not as simply a source for regret. As he puts it at one point, one of the roles of political philosophy is "reconciliation" to our society, our world, and their history. Such a perspective must be developed carefully – else we risk becoming simple apologists for immoral regimes, and whitewash humankind's seemingly unavoidable streak of wickedness.[305] But part of adopting such an attitude responsibly is to recognise that, if we are to value a just society for its own sake, we cannot see the circumstances of justice as giving rise only to problems. They of course do give rise to problems, and unavoidably so. But if the circumstances of justice only gave rise to problems, justice would then only be

301 *TJ*, p. 119/103
302 *TJ*, p. 521/457, *PL*, pp. 201—202
303 See *PL*, pp. 220—222
304 This is clear from many discussions: *TJ*, pp. 5/4—5, 476—477/416—418, 522—529/458—464, 570—572/499—501, *PL*, pp. 147—148, 201—206
305 See *JF*, pp. 3—4. Rawls takes the idea from Hegel: See also *LHMP*, pp. 331—336

remedial, and we would be unequivocally better off if there was no occasion for it. That we live under the circumstances of justice allows us to be just, and this is a matter of celebration, as well as regret.[306]

Having said all this, we can now add to our account of the key first-order interests we are looking to ascribe to the members of the well-ordered society. First, these interests cannot be public first-order interests. Though the members of the well-ordered society value their political institutions intrinsically, those political institutions are designed to fairly meet the various non-public interests of the members of the society. The public institutions of the well-ordered society have similar properties to the capacity for rationality or altruism, i.e. they can be intrinsically valued, but without first-order desires to work on, they have no application. Second, the key interests must also be intrinsic interests. Instrumental interests imply intrinsic interests, after all. In summary then, the key interests of the members are their first-order, non-public, intrinsic interests. We need to ascribe the members of the well-ordered society such interests, otherwise they will not be represented as being motivated to do anything.

To summarise this section: I have stressed that the circumstances of justice are themselves morally neutral, and that they imply a certain degree of divergence of interests, as well as a certain degree of identity. I have mentioned that Rawls conceives this identity of interests to be of both non-public interests, and public interests. The circumstances of justice are not simply taken to be simply a burden on ourselves. They allow valuable ways of life which could not exist in their absence. In the course of describing the circumstances of justice, I have been able to expand the account of the key interests of the members of the well-ordered society, which are now described as first-order, non-public, intrinsic interests.

Section 8: Rawls's Conception of the Person

As was introduced in section 2, and elaborated somewhat in section 4.2 Rawls's conception of the person is of persons as free, equal, rational and reasonable. The account of people's freedom and equality is based upon the account of their reasonableness and rationality.[307] Hence I outline rationality and reasonableness first, in subsection 8.1, and then 8.2. Each discussion will try only to touch on the essentials of the notions, in line with my minimalist ambitions. At the end of 8.2, I outline why these accounts of reasonableness and rationality are not sufficient to ascribe any first-order, non-public, intrinsic interests to

306 For similar reflections on the human condition, see Nussbaum (1990) chapter 15
307 *PL*, p. 19

the members of the well-ordered society. After that, I shall first insert a brief nod to equality (subsection 8.3), and then present Rawls's account of freedom (subsection 8.4). At the close of the latter subsection, I shall again indicate why these two aspects of the persons in the well-ordered society do not ascribe any key first-order interests to those persons.

8.1 Rationality

Rawls's account of rationality has already been touched upon in subsection 4.2. But there, a lot more was left to be discussed.

Rationality is the capacity or power to reflect upon and order our ends, up to the limit of our ends as a whole. Rawls allows that rational deliberation can alter our ends and motivations in ways which go beyond the standard account of means-end rationality. This is explicitly asserted in the later philosophy.[308] In *Theory*, aspects of the discussion suggest means-end restrictions,[309] whilst others do not.[310] But I do not think that there is anything in the book which is in obvious opposition to his later understanding. The key continuities are that rational deliberation always proceeds from our existing motivations, even if these are eventually altered, and that rational deliberation puts no restrictions on what our actual first-order ends might be.[311] Rawls's conception of rationality, then, though not necessarily means-end, is clearly formal.[312]

Rationality consists in the power to form and revise our conception of the good. On occasion, Rawls also refers to our ability to form, revise and pursue a rational plan of life. Our plan of life is our scheme of ends and goals. When organising our plan rationally, we attempt to organise it in accordance with various principles of rational choice.[313] These include the principle of taking effective means to ends — definitive of means/end rationality. They also include (1) the principle to organise our ends so as to ensure that the more inclusive selection of them can be met, (2) to weigh our various final ends by reference to their perceived importance, and (3) to select the more probable over the less probable alternative.[314] Rationality also includes what Rawls calls deliberative rationality: the inspecting and specifying of our ends in order to better understand them and discern

---

308 Ibid. p. 50—51
309 See esp. *TJ*, pp. 415—416/364—365
310 E.g. *TJ*, pp. 412—417/362—367. See in conjunction with *LHMP*, pp. 32—34, 46—47
311 See *TJ*, pp. 432—433/379—380
312 *TJ*, p424/372
313 See *TJ*, p. 407—409/358—359, *PL*, p. 177
314 *TJ*, pp. 411—416/361—365, *PL*, pp. 50, 83, *LHMP*, pp. 32—35, 46—47

their respective weights.[315]

Rawls is clear that a rational plan of life is not necessarily a life of constant deliberation and planning. Rather, the particular individual must simply be contented that they made choices which were overall sensible, and which are not to be regretted, even if they do not turn out for the best.[316]

Rawls employs his account of a rational plan of life in order to give a definition of happiness: "a person is happy when he is in the way of a successful execution (more or less) of a rational plan of life drawn up under (more or less) favourable conditions, and he is reasonably confident that his intentions can be carried through."[317] Under sufficiently unfavourable conditions, even succeeding in a rational plan need not be said to make us happy, as our circumstances (though not our response to them) may simply be too regrettable.[318] Happiness consists in two aspects: the execution of your plan, and the state of mind consisting in the "sure confidence" that your plan will be successful.[319] A person can be happy, moreover, without purposefully pursuing happiness. What we pursue, rather, are the various ends of our plan of life.[320] It might be allowed that they may not even think of this *as* their happiness.[321] This kind of conceptual disagreement is perfectly acceptable, providing that people are able to recognise the importance of the state of the person and attitude which Rawls calls happiness (see below).

Rawls's discussion of happiness segues into a discussion of the possibility of rationally choosing between different rational plans of life. Rawls assesses various traditional solutions to this question which posit a "dominant end" — a single monistic object of value, by reference to which all other values and ends can be subject to arbitration.[322] For Christian philosophers, this is God.[323] For the classical utilitarians, this is

315 *TJ*, pp. 416—424/365—372
316 *TJ*, pp. 422—424/370—372
317 *TJ*, p. 548/480
318 *TJ*, p. 409/359—360
319 *TJ*, p. 549/481. Rawls remarks that happiness can be conceived of objectively and subjectively. In the first, our state of confidence is supported by good reasons. In the second, our state of confidence is based on what we believe are good reasons. But it may be illusionary.
320 *TJ*, p. 550—551/482—483
321 We may again think Rawls in trouble here. Is this *really* how we understand happiness? A full account of happiness will presumably refer to much more extensive facts about human psychology. But we can look on Rawls's account as a minimal account of happiness – or at least an element of happiness – in a similar way that we are currently looking at his account of the person as a minimal account of the moral agent. I say "at least an element of" to leave open the possibility that this may amount to merely a minimal account of something else, such as some attitude of overall contentment with one's life. For recent research into happiness, with one eye firmly on the empirical data, see Haybron (2008) and Tiberius and Plakias (2010).
322 *TJ*, pp. 551—553/484—485
323 *TJ*, pp. 553—554/485—486

pleasurable feeling.[324]

For Rawls, all such theories are examples of teleological theories. His own theory is deontological. He rejects the idea of a dominant end, and instead proposes that rational choice between different rational plans of life is ultimately down to the free choice of the individual agent, providing that choice is constrained by the requirements of a conception of right. Rational choice of a rational plan is, at the last hurdle, down to the standards of the individual agent that they themselves recognise. Hence, with respect to the plurality of individuals, an extremely wide range of quite different plans of life can be rational.[325]

This is the basic idea behind Justice as Fairness's account of what Rawls calls "The Unity of the Self." Rawls speaks of "the unity of the person being manifest in the coherence of his plan" such that "in the ways that justice allows, he is able to formulate and to follow a plan of life and thereby fashion his own unity."[326] These passages should not be taken to be saying too much. All Rawls is saying is that we can be said to have rationally chosen our rational plan of life providing that (1) we have indeed rationally chosen it (i.e. developed it through deliberation and reflection), and (2) it is consistent with the principles of right. There is no need for a further criterion to evaluate plans as more or less rational, such as a dominant end.

Why is the word "unity" used here? The best sense I can make of why it is appropriate is by considering that we can be torn between two or more equally rational plans of life, even after full consideration according to the standards of deliberative rationality. If it ever makes sense to talk of a "self" being dis-unified, it is surely when they are in this kind of predicament. Rawls's answer is to say simply that each person is at liberty in such a situation to decide what will count as a unified character and plan for themselves, given the restrictions of right. The secure institutions and infrastructure of the well-ordered society are assumed to make the ultimate lack of a single criterion for choice less threatening.[327] But it is allowed that the person could choose to remain pulled in both directions, if that is what they truly see as rational. The "unity" of the self, then, does not require the elimination of all tensions and dilemmas in our conception of the good or our commitments.[328] "Unity" seems to have been an extremely misleading word to use here. I cannot recall any of the few discussions I have seen of this section of *Theory* having

---

324 *TJ*, pp. 554—560/486—490
325 See *TJ*, p. 563—567/493—496
326 *TJ*, p. 563—564/493—494
327 *TJ*, p. 563—564/493—494
328 Compare *PL*, p. 44

realised this is all he is talking about.[329] But the central point is that Justice as Fairness has no need to posit something beyond deliberative rationality in virtue of which persons can rationally choose between equally rational ways of life. The account of rationality can hence remain purely formal, and compatible with Rawls's account of freedom.

This leads onto a further aspect of Rawls's perspective of rationality. He attributes certain features to the rational agent, and proposes certain principles they follow. But he does not take himself to be giving a definitive account of the concept of rationality. He is hence proposing a conception of the concept of rationality – admittedly one he does his best to make fairly accommodating. Within certain parameters, he aims to avoid argument with those who have a different conception of rationality. Rawls does not want Justice as Fairness to be unacceptable to those who believe that rational deliberation *is* entirely a matter of means-end reasoning, or who believe rationality incorporates certain ends. With regards the latter example, this is not to admit that Justice as Fairness is compatible with any conception of rationality. Those which incorporate excessively extensive substantive ends into the goals of the rational agent cannot be accommodated within Justice as Fairness (what counts as "excessive" would obviously be a matter of debate). As the above discussion makes clear, conceptions of rationality which incorporate the notion of a dominant end are also at odds with Justice as Fairness.

Rawls believes he can allow a fair amount of latitude, however, as it is important only that his theory can make use of some account(s) of rationality sufficient to complement the account of reasonableness and justice. The account of reasonableness, and more acutely justice, themselves need to be more specific, as they are to arbitrate between the various rational agents. Deciding what is the correct account of rationality is not a moral issue. Deciding on the correct account of justice, by contrast, is.[330] Providing that persons in the well-ordered society can recognise that they are all using acceptable conceptions of rationality, all will be well.[331]

This position of Rawls may be thought to be inconsistent. How can he put forward a conception of rationality, but then maintain that that conception is inessential, and that Justice as Fairness can be accepted even by those who accept some cousin of that conception of rationality? The volume of material Rawls presents on Goodness as Rationality is perhaps excessive given this concession, but that is no reason to think that this general attitude is not acceptable. Rawls is not saying that every conception of

329 Freeman (2007b) pp. 159—161and Sandel (1998) esp. pp. 19—22 are the two discussions I know of. Both appear to introduce more complexity and more substantial claims into Rawls's discussion than are actually there, but I do not argue why I think this here.

330 See *TJ*, pp. 446—447/392—393, 564/494

331 See also *PL*, pp. 176—177

rationality is compatible with Justice as Fairness. Admittedly, he does not precisely elaborate which ones are compatible, but there is no reason to think this task could not be completed

Three final minor points. The accounts of rationality and happiness here, and the rejection of a hedonistic dominant end, should make it clear that happiness for Rawls is not simply a matter of an agreeable or joyful feeling. Happiness is rather the pursuit of and success in our rational plan, and given the connection between rationality and freedom, is also an expression of our freedom.

Rational agents are not always simply individual human beings. Organisations, collectives, companies, etc. can all also be rational agents. Each can have their own distinctive ways of organising their ends into overall schemes, and hence each can have their own variant of rationality, within certain limitations.[332]

Furthermore, rationality is the perspective of an agent's own good. I have hitherto said that the parties in the original position are self-interestedly rational (subsection 4.2). But this is misleading. Self-interest is ambiguous between being only concerned for one's self, i.e. being an egoist, or being only concerned with one's personal interests — interests "of a self" rather than "in a self."[333] Hence our rational interests – our good – can include commitments to our friends, family, community, country, religion, or what have you.[334]

8.2 Reasonableness

Reasonableness is contrasted with, and complements,[335] rationality. When reasonable persons pursue their own good in cooperation with others, they wish to pursue it in a way that is fair to the others they are cooperating with in that pursuit, *and* to themselves. I think of it this way: reasonable persons come to interaction with others unwilling to press for their own good at all costs, and unwilling to use the full powers of their physical or intellectual advantages to get as much as they can. Instead, they stand ready to put aside some of their aims (though not all), providing that others are also willing to put aside theirs, and come to an agreement on fair terms of cooperation.

If reasonable, you approach cooperation with others with your interests viewed as provisional. You are willing to put aside any of these interests, provided that (1) the

332 *PL*, pp. 50, 220—221.
333 *TJ*, p. 127/110
334 *PL*, pp. 50—51
335 *PL*, p. 52. Freeman (2007b) pp. 22—25, tells a Rousseauian story of how neither rationality and reasonableness should be seen to be more primitive than each other. Rawls endorses it at *PL*, p. 53

willingness of others to do the same can be known and assured, and that (2) the cooperation is mutually advantageous to all overall. I believe that *any* interests is the right understanding. A merely reasonable being "would have no ends of their own they wanted to advance by fair cooperation."[336] From this statement I conjecture that an ideally reasonable person would be willing to put aside any of their particular ends, providing it supported a fair and advantageous social world. Only an understanding such as this could be compatible with Rawls's comments in *Theory* that "a perfectly just society should be part of an ideal that rational human beings could desire more than anything else once they have full knowledge and experience of what it was"[337] and "for the sake of justice a man may lose his life where another would live to a later day."[338] This latter example quite starkly supports my reading, I feel. If any interest may be expected not to be put aside, it would be an interest such as this. Of course the sacrifice couldn't be pointless. But then the sacrifice wouldn't be pointless if it was in the service of preserving a just and fair social world.

It may appear that this ideal of reasonableness is excessively demanding. Surely there are certain commitments each of us has which we could never consider giving up in the name of a just and fair social world? What if doing my part in preserving or working towards a just society required I put my family in danger? I was the only one with the information regarding the coup which is being plotted. I know that the coup will most likely succeed unless I act. I also know that revenge will no doubt be pursued if I do act. It is understandable that I may not act — I do not think inaction would make anyone an immoral person here. But there are several observations to make. First, a fair social world would presumably not require that people give up such commitments unless it were completely necessary to maintain the essentials of that society. It will often be the case that other demands can be made of us. These other demands could indeed also be serious sacrifices. But it all comes down to whether failing to do what is necessary to support the fair scheme of cooperation *is* the greater or lesser sacrifice for the person. In my example, my family's safety is a greater sacrifice to me than acting on the ideals of a just society. But placing our home in jeopardy, while a great sacrifice as well, may not be as great a sacrifice as justice. Hence, all-things-considered, acting justly at the cost of the family home may be my only proper response.[339] However, if such choices continually arise, eventually it may be that there is not enough in common between our interests and those of

336 *PL*, p. 52
337 *TJ*, p. 477/418
338 *TJ*, p. 573/502
339 If Mam, Dad or Becky are reading this: first, hi; second, thanks for reading this far; and third, I think justice can let us keep the house.

the rest of society. Cooperation for reciprocal advantage may simply not be possible, and the circumstances of justice will not obtain. So in summary, reasonableness may be very demanding. But it cannot be so demanding that the individual agent actually gains nothing at all from playing their part in a reasonable scheme. This is ruled out by the very definition of reasonableness given by Rawls.

It is important to stress at this point that being motivated to be reasonable is not necessarily to be motivated by some element of one's good. If this were the case, there would be no question as to whether the requirements of justice and right could be congruent with a person's good (subsection 4.2). Either justice would be part of your good, or you would be completely unconcerned with justice. The possibility that you acted justly, but acting justly was not good for you, would not exist. This is not to say that being just is not an end for people who are reasonable. It is simply that it is an open question whether it is a rational end – an open question which congruence arguments attempt to close.

Furthermore, being reasonable – it should be stressed – is not to be motived purely by the elements of other people's good, as with altruism.[340] Rather, reasonable persons

> are not moved by the general good as such but desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept. They insist that reciprocity should hold within that world so that each benefits along with others.[341]

Neither reasonableness, nor rationality, nor indeed altruism, are sufficient in themselves to ascribe much character to the members of the well-ordered society. As noted earlier, interests relating to each of these capacities are second-order interests, i.e. interests about one's other interests, or the interests of others. This is obvious with altruism. It moves us to meet others' interests, but not our own. But the other powers presuppose first-order motivations and interests also. Rationality entreats us to organise our various ends and motivations. But need not itself, on Rawls's view, provide any. Reasonableness asks us to secure arrangements between oneself and others where a fair selection of everyone's interests are met. But, again, there must first be initial interests to balance. Having said that the members of the well-ordered society are reasonable and rational, we hold that whatever interests and ends they possess, they will order them and attempt to act on them within the limits set by rationality and reasonableness. But we have not said anything about what

340 *PL*, p. 50, *TJ*, p. 189/165
341 *PL*, p. 50. See also *TJ*, p. 478/418—419

those interests etc. will actually be.[342] Included amongst these interests are the key interests previously spoken of. Until we have said something about these key interests, we cannot say that the members of the society will actually *do* anything at all.

8.3 Equality

The members of the well-ordered society are equal due to "their having [the moral] powers [or the capacity for them] to the requisite minimum degree to be fully cooperating members of society."[343] The criterion for equality will be discussed at length in the final chapter. Hence I do not further discuss equality here.

8.4 Freedom

It is

> In virtue of their two moral powers (a capacity for a sense of justice and for a conception of the good) and the powers of reason (of judgement, thought, and inference connected with these powers), [that] persons are free.[344]

As outlined in *Political Liberalism,* and *Justice as Fairness: A Restatement,* a person's freedom has three aspects.[345]

(i) Persons are free in that they are capable of developing and revising their conception of the good. Persons do not see themselves as tied to any of their particular ends. They are able to reflectively appraise them and decide whether they affirm them.[346] Being free in this way follows directly from the members of the well-ordered society having the moral power of rationality.

(ii) Persons are free in that they regard themselves as "self-authenticating sources of valid claims". They take themselves to be able to make claims on their shared

342 See further *PL*, p. 48 fn1
343 *PL*, p. 19
344 *PL*, p. 19
345 I do not investigate how these accounts given in the later philosophy are foreshadowed in *A Theory of Justice* and its immediately subsequent articles. But I believe that they are so foreshadowed, and that they do not represent drastic departures from anything found there. They are first introduced in this form in "Kantian Constructivism in Moral Theory" (*CP*, chapter 16)
346 *PL*, pp. 30—32, *JF*, pp. 21—22. See also *TJ*, pp. /131—132, 408/358—359, 416/365—366, 561/491—492

institutions. In regarding themselves as able to authenticate those claims themselves, they do not regard themselves as only able to make claims in virtue of prior "duties and obligations owed to society".[347]

Though Rawls is not explicit, I believe that this aspect of freedom is best seen as arising from the fact that persons are, and view themselves as, reasonable. Reasonableness considers not just others apart from the agent, but the agent themselves as well, and the relationship between the agent and others. Each must be seen, and must see themselves, as a source of valid claims in order for a fair and mutually advantageous arrangement to be generated.

By contrast, rationality appears to be perfectly compatible with not viewing ourselves as the ultimate sources of the claims we make. As an example of those who do not see any claims they may make as ultimately originating from themselves, but from others in their society, Rawls cites slaves who have completely internalised the way they are regarded by a slave-owning polis. It is arguable that such slaves could still be described as rational.[348] The idea would be that they can sensibly organise, pursue, and perhaps even to a certain extent adapt the ends which another has ascribed to them.[349] But such slaves cannot act reasonably so long as they fail to recognise themselves as a sovereign agent as well.

(iii) Persons are free in that they understand themselves to be responsible for their ends, given the just institutions of their society. They do not take the simple *strength* of any of their desires on its own to constitute a reason for society to fulfil that desire, or for them to act as they can in order to meet that desire themselves.[350] Rather, reasons are based on the *authority* which a desire possesses, due to its being endorsed as reasonable, and, perhaps, rational.[351]

This aspect of freedom may be seen to derive from persons' reasonableness and rationality. Rationality allows us to separate out the strength of our desires from their authority, as we can view a powerful desire which we acted on as nevertheless against our own rational interest. Reasonableness furthermore leads us not to make demands on our society which go beyond what is fair, even if we strongly desire that such demands should

---

347 *PL*, pp. 32—33, *JF*, pp. 23—24

348 Dudley Knowles has urged me to remember that some would disagree with this – Hegel most prominently. I think that the position I give here, however, is likely to be Rawls's position, given how he understands rationality.

349 Given how these notions have so far been formally defined, the fully compliant slave appears to have an attitude equivalent to *altruism*. Even if conceived like this however, rationality can still be ascribed to the slave. Altruistic interests, like any others, can be pursued rationally or irrationally.

350 *PL*, pp. 33—34, 185—187

351 See *PL*, pp. 82—86

be met, or that we should make such demands.

Once again, neither freedom nor equality are sufficient by themselves to ascribe any first-order interests to the members of the well-ordered society. Simply saying that people regard each other as having equal possession of the moral powers does not suffice to ascribe them any aims, as the interests connected to those powers are all second-order. Regarding the aspects of freedom: saying (i) that the members of the well-ordered society do not see themselves as unavoidably tied to any of their particular aims does not tell us anything about what those aims could be. Seeing one's claims as having some kind of authority independent of what society demands of you – (ii) – and seeing yourself as responsible for whatever claims you make, providing your society is reasonably well-ordered, and that you cannot reasonably claim whatever you might desire – (iii) – similarly do not determine what the claims and desires in question are.

In particular, regarding (ii), note that Rawls's account does not commit him to the position that people in the well-ordered society will see themselves as a self-authenticating source of valid claims *per se*. This would allow us to say that their first-order interests include interests in themselves, i.e. in their own well-being. But this is not so. Rather, the persons in the well-ordered society see themselves as a self-interested source of valid claims with respect to the rest of their society as a whole. According to their own understanding of their first-order interests, these interests may not be conceived as primarily *their* interests at all. They may be of a largely altruistic cast of mind, and see their interests as only instrumental for others interests — such as those of their family, or club, or ethnic group. Whole groups of people – even whole societies – may have this kind of mindset, if they simply see their interests as ultimately based on the interests of fictional, transcendent, or non-human being or beings. I may view my interests as validated *solely* because they serve the interests of God, the Animals, Nature, or the Justified Ancients of Mu Mu. Of course, some of these are outlandish possibilities, given human nature (we might think). But formally, nothing in Rawls's account of the conception of the person rules them out. That these possibilities are compatible with his account of freedom is clearer in Rawls's later philosophy.[352] But they follow from the earlier account as well.[353]

Finally, we might assert that the conception of the person is as yet incomplete.

352 See, for example, *PL* pp. 32—33

353 E.g. *TJ*, pp. 127/110. I believe that several aspects of Rawls's account of moral development, and discussions of self—respect and self—esteem, do not properly respect this fact (e.g. *TJ*, pp. 463—465/406—408, 498—501/436—439). This may be one reason why the original account of Justice as Fairness needed to be revised (see subsections 12.1, 12.2). But I do not explore this issue here.

Though we have described the two moral powers, and the nature of the equality and freedom that follows from them, it must be remembered that the conception of the person is of a person living their life in a well-ordered society. We must be able to ascribe first-order, non-public, intrinsic interests to such persons – a person can hardly be described to be living their life if they have no interests or desires which actually spur them to action. The picture will be completed in the next section.

To summarise this section: all the characteristics so far attributed to the members of the well-ordered society merely serve to narrow down the possibilities of the first-order interests the members of the society. So far, the members will ideally not pursue their interests and attachments in ways which are obviously irrational.[354] They may have interests which would be unreasonable to claim, but they will not press for those interests to be met. They will consider themselves as equal in possessing the powers to be able to act this way. And they will consider themselves to be free of being tied to a particular set of commitments, to be free to claim the authority of their own claims, and will see their claims as outcomes of their free agency, and hence as their responsibility. They will have an understanding of what their happiness is. This all narrows the range of possible conceptions of the good — of possible systems of ends. But we have still specified nothing positive about the first-order, non-public, intrinsic interests of the members of the well-ordered society. Without such interests, none of these second order interests, or attitudes towards our interests, will have any application.

Section 9: The Conception of the Person and Human Psychology

What I have attempted to show in the preceding discussion is that Rawls's account of the person in the well-ordered society does not tell us much at all about the character of such persons. In particular, it does not give such a person's key first-order interests. For these various character traits and interests are all second-order. When, then, do we acquire the information about the key interests of the members of the society? I believe that the best answer is to be found from considering the circumstances of justice. I defend the idea that the circumstances of justice and the conception of the person combined are sufficient to ascribe key first-order, non-public, intrinsic interests to the members of the well-ordered society, and that we need not have to recourse to postulating more specific facts about persons' psychologies. Once we have this, I believe we will be able to delineate the

354 "Obviously" because, as stated, the concept of rationality is open to wide interpretation.

minimal core of the normative conception of the person found in Rawls's theory. This all happens in subsection 9.1 below. From this basis, we are then well placed to clearly distinguish the various further features which Rawls attributes to the members of the well-ordered society. We can see these as primarily arising from psychological claims, now that we have the normative essentials of his position. This is not to say that they exclude normative claims themselves, as was noted in subsection 6.1 above. But these latter claims will have a different standing in the theory, as they will be based on Rawls's assumptions about human nature in the way the conception of the person is not. Such will be the subject of subsection 9.2.

9.1 Key interests and the circumstances of justice

This section will first outline how the conception of the person, combined with the circumstances of justice, are sufficient to ascribe first-order, non-public interests to the members of the well-ordered society. It will then argue that the specification of the psychological, biological, and sociological attributes required for human beings to realise the moral powers – which are also ascribed to the members of the well-ordered society – do not necessarily specify any first-order, non-public ends to the members of the society. Similarly, the specification of the psychological, biological and sociological attributes which are provided to each member of the society as part of the social primary goods does not, necessarily, ascribe any first-order, non-public ends to the members of the society. Such first-order ends are only necessarily ascribed if the correct account of human nature indicates that they would have to be present in order for the well-ordered society to be sufficiently stable.

We have been searching for the non-public, intrinsic first-order interests of the members of the well-ordered society. But we have yet to find them. None fall out of the conception of the person Rawls offers. However, I believe that we can attribute first-order ends of the members of the well-ordered society simply by considering how the circumstances of justice relate to the conception of the person. I first describe how the circumstances of justice and the conception of the person together can be understood to give sufficient information about the parties in order to say that they have suitable first-order ends. I then answer some possible misgivings those familiar with Rawls's texts will have about the answer I give here.

I ask you to recall that the subjective circumstances of justice specify that human beings' interests and ends are such that social cooperation is mutually beneficial, though it

still allows and gives rise to many conflicts of interest. Purely on the basis of this, we can say that the first-order interests of the members of the well-ordered society are such that social cooperation is beneficial to the members, i.e. they have interests which they can meet only through cooperation. It also allows us to say that they have first-order interests which are in conflict. These two sets of interests are not mutually exclusive. I may have an interest which I can only meet through social cooperation, but which is in conflict with others' interests. It cannot be said in advance whether any such interest will be met for that person, as was indicated in the discussion of reasonableness above. What can be said is that a fair number of every person's interests will be met. If this were not the case, then the society would simply not *be* a fair system of cooperation. Furthermore, it will be remembered that the interests which are met include ends for which the political order of the well-ordered society is instrumental, and ends in virtue of which the political order is intrinsically valuable. However, it is the case that the political order rests on their being non-public interests which need to be fairly governed by that order. This is the task of political institutions, so if there are no non-public, first-order interests to fairly adjudicate, then political institutions have no such task to perform. Hence, though public ends can be first-order ends, like the first-order interests in being altruistic considered in subsection 6.2 above, they are not the right sort of first-order interests we need to allow us to ascribe first-order interests to the members of the well-ordered society. The ascription of non-public, first-order interests is essential.[355] Finally, non-public, first-order interests also include intrinsic interests.

All this follows from the circumstances of justice. But, as was noted in subsection 7, the circumstances of justice themselves do not get us justice. Indeed, they do not even get us social cooperation. That is why the account of the circumstances of justice needs to be combined with the account of the person with two moral powers. It is by the exercise of these powers that the interests given by the circumstances of justice can be met. In order to get social cooperation, a rough answer would be that we need the circumstances of justice,

355 It might be thought that these comments about persons having ends which can only be met through social cooperation, and persons having ends which are nevertheless in conflict, only applies to non-public ends. I do not think this need not be the case. There can be dispute about the precise way that public institutions are arranged, and about the precise shape of the public culture. Different groups and persons can prefer different arrangements. What is again required is, again, that a fair number of each of these public interests will be met. There are complications looming here. These disputes about the arrangement of public institutions must presumably stop at some point, as everyone will need to agree to certain institutions whose job is to arbitrate between the rest of the arrangements of society. I have in mind here elements of a society's constitution (*TJ*, pp. 195—196/171—172). There must also be sufficient shared content in the public conception of justice to allow acts of civil disobedience to appeal to a shared conception of justice (*TJ*, pp. 365—366). But I leave this matter and others aside here.

and rational agents.[356] In order to get just social cooperation, we need in addition agents who are reasonable. If agents were not reasonable and rational in these ways, a scheme of social cooperation would never even get off the ground. The full, minimal conception of the person is constituted by the combination of the circumstances of justice (from which we get first-order, public and non-public interests) and the moral powers (from which we get the various second-order interests).

We have finally found some first-order, non-public interests to ascribe to the members of the well-ordered society. They are given by the assumption that the circumstances of justice obtain, and that persons are reasonable and rational. We can now present the most fundamental elements of Rawls's conception of the person. Persons are free, equal, rational and reasonable, which means that their interests will be organised in certain ways. Such persons find political cooperation beneficial to their interests, both instrumentally and intrinsically, but also find that some of their interests conflict.

This specification of the first-order interests of the members of the well-ordered society tells us very little about the content of the interests themselves. Indeed, all it says is that they have first-order interests which can be met only through cooperation with each other, and also interests which are in conflict. We are only to make further specifications of the member's first-order interests in view of the general psychological, biological and sociological facts about human beings which the parties have access to in the original position.

These elements represent a significant chunk of the normative fundamentals of Justice as Fairness (others may arise from the conception of the well-ordered society, see section 11 — but also see section 15, particularly subsection 15.5 L). Any further elements of Justice as Fairness are developed in the light of the facts of human nature as they are considered by the parties in the original position. Aside from these facts, the conception of the person (combined with the conception of the well-ordered society) would be appropriate to form the foundation for the derivation of principles for *any* type of reasonable and rational agent, whether human or not. Justice as Fairness, and its distinctive principles of Right and ideas of the Good, is developed in the light of human nature. As Rawls writes: "justice as fairness is a theory of human justice and among its premises are the elementary facts about persons and their place in nature."[357]

It may be thought that this interpretation runs against what Rawls actually says. It

356 I say rough answer, as remarks about the inherently reasonable (to some minimal extent) nature of social cooperation at *PL*, p. 16 complicate matters. Presumably, ideal reasonableness yields justice, but human beings can be reasonable but less than ideally reasonable, and hence merely socially cooperate. I ignore these complications here.

357 *TJ*, p. 257/226. See also pp. 159—161/137—139

might be thought that Rawls builds more robust assumptions about human nature directly into the specification of his conception of the person and their first-order ends. In *A Theory of Justice*, having given his account of deliberative rationality, Rawls then remarks that the account remains purely formal. In order to develop an account of what he calls the social primary goods (see subsection 15.2), so as to derive principles of justice, Rawls makes clear that we now have to attend to "certain general facts" about human nature. These include "the broad features of human desires and needs, their relative urgency and cycles of recurrence, and their phases of development as affected by physiological and other circumstances" and "the requirements of human capacities and abilities, their trends of maturation and growth, and how they are best trained and educated for this or that purpose."[358] It might be thought that reference to these facts may specify the content of some of the non-public, first-order, intrinsic interests of the members of the well-ordered society. Well they may. But it is completely unnecessary to assert this when outlining the conception of the person. There only arises a need to assert this if the parties in the original position, considering the facts of human nature, discover that we need to specify that possession of such characteristics must be amongst the first-order, non-public, intrinsic interests of the members of the well-ordered society in order for the society to be stable.

I shall first consider whether the psychological and biological needs which underlie the development of the two moral powers ascribe any non-public, first-order, intrinsic ends to the members of the well-ordered society. I shall conclude no. I shall then consider whether the presuppositions about human nature made in specifying the social primary goods ascribe any such ends either. Again, my answer will be no.

The moral powers are capacities which are embodied by the members of the well-ordered society. Given human nature and our environment, there are certain basic needs – biological, psychological, physical, social etc. – which will need to be met whenever the moral powers are to be realised.[359] Physical health, and the absence of physical pain, for example, are generally needed if we are to be able to adequately exercise our two moral powers.[360]

It is likely that the members of the well-ordered society will have first-order, non-public interests in having these attributes. They are the sorts of things often valued in their own right, and not simply as instrumental for the realisation of the moral powers. But we cannot assume this for the persons in the well-ordered society simply on the basis of the conception of the person. We are only able to say that the members of the well-ordered

358 *TJ*, p. 424/372—373. See also *PL*, p. 178
359 See *PL*, pp. 177 and 178
360 As suggested at *PL*, pp. 181—182

society have these biological and psychological characteristics, and inhabit these kinds of physical and social surroundings, because human nature is such that they are required in order to develop the moral powers. But even if this is true, this does not imply that the members of the well-ordered society value these attributes in themselves. And we only have reason to explore whether they value these things in themselves to the extent that this is relevant to the parties' considerations in the original position regarding the stability of different conceptions of justice. We can say the members of the society value these biological and psychological considerations instrumentally, because they are assumed to value the moral powers. But we cannot say more than that, unless stability considerations turn out to require more to be said.

Rawls also holds that the members of the well-ordered society will have access to a fair allotment of social primary goods. These are to allow each to follow their conception of the good. Such social primary goods will allow the realisation of the psychological and physical attributes which allow those conceptions of the good to be pursued. It is assumed that there is enough overlap in permissible conceptions of the good such that we can specify certain basic human needs – again biological, physical, psychological, social etc. – which are required for the pursuit of any permissible conception of the good.[361]

Once again, however, this stipulation by itself does not tell us what the first-order, non-public, intrinsic interests of the members of the well-ordered society will be. It is assumed that there is a set of basic needs which must be met, and means which must be provided, for the pursuit of any permissible conception of the good in the well-ordered society. But there is no reason to assume that the social primary goods held by the members of the well-ordered society are valued intrinsically by the members of the well-ordered society. Or rather, no need to assume this simply on the basis of the conception of the person, aside from considerations of what human nature appears to demand for stability.[362] Indeed, Rawls holds that these basic needs, and the goods which meet them, are to be thought of as instrumental for meeting our conceptions of the good.[363] Again, it is likely that they are often intrinsically valued. But we need some grounds for saying this, other than that they are means which help any conception of the good to be met.

It should be noted here that there may exist scepticism about whether any workable

361 See *PL*, pp. 176—178, 180 fn8, *TJ*, pp. 424—425/372—373, 433—434/380—381

362 It is true that interests in the social primary goods are first-order and not second-order interests. This is because they are interests in what everyone wants regardless of what else they want. And the social primary goods are also, often, non-public interests. However, they need not be intrinsic interests, which is the significant thing.

363 E.g. *TJ*, p. 93/

conception of the social primary goods can be developed.[364] Are there resources such that any conception of the good can make use of them? As an extreme example, what about ascetic conceptions of the good, which deny any value to material possessions? This problem is avoided – as far as it can be – by the formal features of the circumstances of justice. It is simply stipulated that the members of a well-ordered society have interests which can only be met through social cooperation. Given the assumption that they want to cooperate, and cooperate fairly, they will be able to decide amongst themselves what is required for each one to meet the interests they can fairly advance. Of course, the question then arises whether it is necessarily the case that every individual who finds themselves in the well-ordered society will feel they have anything to gain by cooperating in its institutions. I see no reason to presuppose so. What would then need to be investigated is what can be asked of such a person, or group of persons, by the members of the well-ordered society, and more specifically what (if anything) can be reasonably asked. This is a big topic, and I leave it aside here.

Hence, I hold that we can ascribe first-order, non-public, intrinsic interests to the members of the well-ordered society of Justice as Fairness simply by stipulating that they all can benefit, in a fair manner, from social and political cooperation, and that they are rational and reasonable persons.

9.2 Psychological facts in the original position

I have delineated the essential features of the members of the well-ordered society which can be ascribed to them on the basis of the normative conception of the person employed by Rawls. Such persons are reasonable, rational, free and equal, and possess a conception of the good, which can be furthered by social cooperation, but which is also in partial conflict with the interests of the other free, equal, rational and reasonable persons in their society. Any further things we say about the psychology and character of the members of the well-ordered society follow not from this conception of the person, but from the facts about human nature which are made available to the parties in the original position. To argue for Justice as Fairness and its principles, Rawls makes various claims about these psychological facts, and these give the particular character of his version of the well-ordered society.

364 Such scepticism was expressed early after the publication of *A Theory of Justice* by Nagel (1975) and Schwartz (1973). Rawls's response to Nagel can be found at *PL*, p. 196 fn31, and the surrounding passages.

At this point, let's remind ourselves of the place and roles of moral psychology within the original position, once again. From section 4 we have:

-Psychological facts have the role of showing which conceptions of justice avoid futility, and hence are realisable and stable over time in favourable conditions. Psychological facts also have the role of showing which conception of justice is more likely to be stable than its rivals.

-Psychological facts are employed in both parts of the argument from the original position, with different readings being available for when and how exactly they are so employed. The two extreme readings or renderings, situated at opposite poles to each other, are these. Psychology could not play an arbitrating role at all in the first part of the argument. Hence any conception which was stable could be chosen there, and would only need to be shown to be non-futile in the second part of the argument. Or arbitration between different full moral psychologies of different conceptions of justice could occur in the first part of the argument. In between these two options, there exists a continuum of others.

From subsection 5.2, and section 11 (which is yet to come) we have

- Once the principles of justice have been provisionally chosen in the first part of the argument, they can then be referred to in developing a moral psychology corresponding to those principles. Hence moral psychology presupposes moral principles. Developing a moral psychology also plays a role in further specifying moral principles. The principles which are chosen at the end of the second part of the argument from the original position are the fully justified principles of justice, and are the principles which have normative authority (this will be emphasised in section 11).

- In saying that psychology helps to specify moral principles, but that our particular moral psychology depends upon assumed moral principles, we also seem warranted in saying similar things about the relationship between the conception of the person and moral psychology. So developing our moral psychology presupposes a certain conception of the person. But in the course of developing our moral psychology we can also specify our conception of the person. However, unlike with principles of justice, the basic (minimal) conception of the person remains unchanged throughout this (as was stated in subsection 6.1).

From this, we can draw the following distinctions. We can distinguish between those aspects of human psychology which are *constitutive* of the conception of the person when it is embodied in human beings, and those aspects of human psychology which are *supportive* of the conception of the person being embodied in human beings. The first of these is composed from what we might call the minimal content of the conception of the person, which we have outlined above, and also the content which is added to that conception of the person when we come to specify that conception of the person with regards to the case of human beings. The second of these is composed of the necessary and sufficient psychological and biological characteristics which a person must have in order to be able to realise the conception of the person, but which are not themselves elements of that conception of the person, and hence cannot be said to be constitutive of it.

The relationship between the conception of the person, between a set of principles of justice which is in accordance with that conception of the person, and between human psychology overall, appears to be this. Part of the content of any viable set of principles of justice must be the provision of the resources needed for human beings to realise the conception of the person — i.e. to realise the two moral powers and meet some minimal level of a fair determinate conception of the good. This consists in the provision of resources which are sufficient to allow a human being's capacity to realise the conception of the person to indeed be realised — both the resources which constitute that capacity, and those which support it. The further content of any viable set of principles of justice will detail how the further goods of social cooperation are going to be distributed, i.e. how social primary goods are going to be distributed fairly between all the different moral persons. In other words, the further content of any set of principles of justice distributes the social primary goods which are surplus to realising everyone's moral powers, and meeting everyone's conceptions of the good to a minimal level (this latter minimum must be met for everyone in order for social cooperation to be mutually beneficial, and hence for the society to even be a well-ordered society).[365]

365 Rawls describes what this minimum must be clearest at *JF*, pp. 97—100 esp. fn21. The minimum is basically the equivalent in Rawls's theory to the non-cooperation point in social contract theory generally. What this non-cooperation point is varies from theory to theory. For example, in Hobbes (1651) it is the State of Nature (see the famous chpt 13). Rawls's theory is distinctive in that his "non-cooperation point" is actually quite high. In fact, it is not really accurate to describe it as a non-cooperation, given that he believes that cooperation can exist which is not just (see section 7 above). Rawls himself describes it as the "guaranteeable level" which allows for at least a "satisfactory political and social world" (*JF*, pp. 99—100, see also pp. 127—130). It might alternatively be characterised as the minimally just, or even better minimally reasonable cooperation point. Rawls believes that Justice as Fairness is the most reasonable conception of justice as it represents the fairest departure from the minimally reasonable cooperation point. On non-

In order to be able to come to a decision in the original position, given the fundamental interests of their representees in being moral persons, then, the parties there must be able to determine the following, through reflection on human psychology. They must determine what is *essential* for it to be non-futile that (1) human beings realise their moral powers and that (2) human beings each receive a fair allotment of the social primary goods. Furthermore, given this, the parties must also consider, given the facts of human nature, what makes it *more likely* for (1) and (2) to obtain. Finally, they must also consider which principles allow the fairest allotment of social primary goods, given the previous two requirements have been considered.

Departing from the minimal account of the conception of the person, we can try to account for the presence of the other features of the members of the well-ordered society and their principles of justice as specified by Rawls, and what they imply about his assumptions of human nature. Whereas Rawls worked outwards from a conception of the person, through the various stages of the argument from the original position, by reference to human nature, to a set of principles of justice, we can take the conception of the person, the set of principles, and the set-up of the original position as given, and attempt to reconstruct in more careful detail the assumptions about human nature which Rawls makes, both explicitly and implicitly.

As noted above, and in the introduction, I shall not be attempting this task here. I know for a fact that it would be a serious labour. Such an account would first have to outline the psychology presupposed by Rawls's accounts of both moral powers — the ability to develop a conception of the good, and the sense of justice. As noted in subsection 3.1 above, we should view all these, plus the presence of a determinate conception of the good, as elements of the moral psychology of the members of the well-ordered society. The study would then need to take in the accounts of self-respect and self-esteem — a topic which has been well covered in itself, but which has still not been linked up to the rest of his moral psychology in a systematic way.[366] There would also been a need to look in depth at various more particular psychological posits, such as the Aristotelian Principle.[367] Beyond the account of self respect, there would be a need for a thorough assessment of Rawls's argument for the primary goods from a psychological perspective. The account of moral development is complex in its detail, and has nowhere, I feel, been

---

cooperation points (or as they are sometimes called, non-cooperative baselines, or nonagreement points) in general, see Barry (1989) esp. chapter II.

366 See *TJ*, pp. 440—446/386—391, *PL*, pp. 318—319. Recent work includes Zaino (1998), Eyal (2005), Doppelt (2009), Zink (2011)

367 See *TJ*, pp. 424—433/372—380, 440—441/386—387, 528/463, 571—572/500—501, *PL*, pp. 203 fn35, 207

adequately covered. Nor has it been explored how compatible it still is with subsequent developments in the field, or with Rawls's own modifications of his theory.[368] An account of moral development will need to be linked to an account of the moral emotions, and their relations to the moral and natural sentiments.[369] The account of the moral motive in Rawls – the motivation to be a just and reasonable person — has not been systematically explored, nor the more general notion of acting on principles of right, as opposed to habit or custom.[370] The notion of a psychological reciprocity principle will need to be examined at length.[371] The discussion of envy, and the other special psychologies, need to be related to related to what has gone before.[372] And I feel we still lack an adequate treatment of the congruence argument in all its details.[373]

In both earlier and in subsequent chapters, I have made occasional references to the content of this moral psychology when appropriate for my arguments or outlining my position. But I do not view these scattered comments as at all amounting to a full account of Rawls's moral psychology, though I have tried to fit them as best I can to what my current understanding of the overall shape of that psychology currently is. But, as I have said before, a thorough account of Rawls's moral psychology would take a whole other thesis in itself.

368 See in particular *TJ*, pp. 467—479/409—419. Existing discussions include Pritchard (1977), Kearns (1983), Alford (1991) chpt 7, and Baldwin (2008) pp. 258—261

369 See *TJ*, pp. 442—446/388—391, 479—490/420—429. Important discussions include Deigh (1982), Taylor (1985) chpts III and IV

370 See *TJ*, pp. 476—479/416—419, *PL*, pp. 48—54, 82—86. Discussions include Bates (1974), Barry (1995b), Freeman (2003), (2007b) chpt 5. See also Schwarzenbach (2009) pp. 82—88

371 See *TJ*, pp. 490—496/429—434. Pritchard (1977) and Ci (2006) chapter 7 discuss this

372 *TJ*, pp. 530—541/464—474. This is often covered in the literature on self-esteem and self-respect. See, in particular, Zaino (1998)

373 See *TJ*, pp. 567—577/496—505, *PL*, pp. 201—206, *JF*, pp. 198—202. For discussion, see Barry (1995b), Freeman (2003), (2007b) chapters 5 and 6, Mendus, chapter 4

Chapter 5: Moral Psychology in Political Liberalism

This chapter explores the connection between Rawls's development of his political liberalism and the question of stability. It also considers, as a distinct issue, what relevance Rawls's shift to his later philosophy has to the roles and content of his moral psychology, though it comes to no definite conclusions as regards the content. Section 10 presents an outline of the chapter. Section 11 introduces and examines the notion of stability for the right reasons, which led to Rawls's adopting the idea of political liberalism. Section 12, after first outlining the major features of political liberalism, then reconstructs why Rawls's commitment to stability for the right reasons led to political liberalism. In section 13, I then remark on alterations in Rawls's use of his moral psychology during the politically liberal period. I highlight how the roles of moral psychology are essentially the same in the later philosophy as in the earlier philosophy. I then give some small indication of the extent to which the shift to political liberalism could influence the content of Rawls's moral psychology.

Section 10: Outline of the Chapter

The initial focus of this chapter is not on moral psychology. Its initial concern, rather, is to try to get clear on the reasons why the question of stability led to Rawls revising his theory. I believe that only after getting clear on this we can begin to examine to what extent Rawls's later political liberalism forces changes to his moral psychology. I do not explore extensively what alterations may be needed. Once again, as in chapter 4, I believe that more extensive empirical research is needed. But, in addition, my analysis will leave the more precise relationship between psychology and justification in the original position in the later philosophy for the most part unexamined. Hence, even more so than my other chapters, this chapter only represents a groundwork for assessing the content of Justice as Fairness's moral psychology.

Other discussions of why the concern with stability led to the revisions of *Political Liberalism* have proceeded differently. They have looked to individual elements of the stability argument as it was presented in *Theory*, with the aim of determining which of them required Rawls to revise his views.[374] I believe this is a less than ideal approach – though I would not claim it is hopeless. It may be that there are elements of the stability

374 At least in part, this is the approach adopted by Barry (1995b), and Freeman (2003), (2007b) chapters 5, 6.

argument and of the moral psychology which need to be dropped, in order for the stability argument to succeed, given the further assumptions Rawls adds to his theory in his later period. Against most critical opinion, I actually doubt that this is the case. I suspect that Rawls actually retains all of his existing arguments in some altered form. I believe that he can do so, providing he can argue, amongst other things, that the account of human nature in *Theory* is still correct in the appropriate ways. But even if Rawls is wrong in his account of human nature, this would not make the approach I am pitting myself against here the right approach. We should first seek to attain a clear understanding of just what the additional premises were which Rawls added to his theory, and from these reconstruct why the changes were necessary. From this foundation, we can then accurately assess what changes Rawls did make, and even more importantly what changes he should have made.

As mentioned, the approach I adopt shares some commonality with the approach in the previous chapter. I attempt to find the most basic and minimal presuppositions which Rawls must have made in order for him to judge that his theory needed to be revised. Once we have these, then we can investigate why these impact on the other elements of the theory in the way they do. The overall aim of the chapter, then, is to investigate what changes Rawls's political liberalism brings for the content and roles of moral psychology within his theory, *by first* developing an understanding of why considerations of stability led Rawls to revise his theory in the first place. The chapter will proceed by first examining this latter issue (sections 11 and 12). I then move on to the former issue, though I consider it more briefly (section 13).

I first clarify the notion of stability which is at work in Rawls (section 11). I then introduce the basic ideas of political liberalism (section 12). I then turn to Rawls's account of why stability forced changes in Justice as Fairness. On investigation, Rawls's reasons for the changes he made are presented by him in a misleading fashion.

I will then turn to what ramifications this had for the moral psychology. I first outline how Rawls's moral psychology is employed in the later work – highlighting that though it is employed slightly more expansively than in the earlier work, it plays basically the same roles. I then very briefly comment on what sort of impact restrictions from the idea of political liberalism might make on the content of the moral psychology, when it is brought forward from the earlier work. I argue this could well be quite slight, but aim to leave the matter open, and to urge caution. All of this occurs over section 13.

Section 11: Stability for the Right Reasons

What led Rawls to make the changes he did? I have so far said that Rawls holds that any conception of justice, to be fully justified, must be able to be accompanied by a moral psychology which shows how the well-ordered society corresponding to that conception can be stable over time (subsections 3.2, 3.3, section 4). But the notion of stability has not been presented with all its elements as of yet (except in summary fashion in subsection 3.2. It was also briefly mentioned again in subsection 5.2). Rawls's full commitment is that any theory of a well-ordered society must show how that society can be "stable for the right reasons."[375]

In this section, I first remark on when Rawls introduced the phrase, though not the idea, of stability for the right reasons to his theory. I then outline the two elements of stability for the right reasons, and I then show how they are interdependent.

Rawls does not make it easy to recognise just when and where the idea of stability for the right reasons is at work in his theory. The phrase "stability for the right reasons" was only introduced in "Reply to Habermas",[376] and then subsequently incorporated into the introduction to the paperback edition of *Political Liberalism*. There Rawls remarks that

> The phrase "stability for the right reasons" does not occur in the text of *PL*, but "stability" should usually be given that meaning in both *Theory* and *PL*, as the context determines.[377]

The idea of stability for the right reasons is present in the earlier and the later philosophy. To illustrate this continuity with an example, when discussing "the criterion of stability" in *Theory*, Rawls notes that "some ethical theories have flouted it entirely." The example he gives is of an interpretation of Benthamite Utilitarianism, in which psychological egoism is presumed. The utilitarian legislator arranges society's institutions so that, nevertheless, an artificial identification of interests results.[378] But surely such a society could be stable and persist over time? The point is that it would not be the kind of stability Rawls is interested in. Rawls's contrast can only make sense if the meaning behind his use of the word "stability" in this section of *Theory* is stability for the right reasons.[379] Stability for the right reasons, then, does not enter only with Rawls's political liberalism.

Stability for the right reasons is a characteristic of well-ordered societies, and it has two elements. Stability obtains for the *right reasons* when the society is governed by a

375 *PL*, p. xxxvii
376 *PL*, pp. 388 fn21, 390, 392
377 *PL*, p. xxxvii
378 *TJ*, p. 455/399
379 *TJ*, pp. 453—455/397—399. For confirmation, see *PL*, p. xl

public conception of justice (section 2). The institutional arrangements of the society are hence publicly justified, and the society meets what Rawls calls the liberal principle of legitimacy (see below). Such a society, organised in accordance with the right sorts of reasons, is *stable* when human beings growing up in such a society are liable to develop a sense of justice strong enough to lead them to act so as to support the basic institutions of that society (subsection 3.2). Both these elements are essential. Not only must the "character and interests [people form] by living under a just basic structure [be] strong enough to resist the normal tendencies to injustice." In addition, the people's support of the justice of the society cannot be merely a function of them acting on the basis of other reasons or motivations.[380] It must based on their express and "reasoned" support.[381] In the manner I put it earlier in subsection 3.2, it must be in virtue of the members of the society being moved by the reasons found in the public conception that the well-ordered society and its institutions are sustained.

I now discuss each half of the idea of stability for the right reasons in turn. The second half of the "stability for the right reasons" slogan connects to the widely (though not ubiquitously)[382] acknowledged liberal ideal of public justification. Jeremy Waldron eloquently characterises it:

> the social order must be one that can be justified to the people who have to live under it ... a *transparent* order, in the sense that its workings and principles should be well-known and available for public apprehension and scrutiny. People should know and understand the reasons for the basic distribution of wealth, power, authority, and freedom.[383]

This idea is not unique to Rawls, and is not restricted to his political liberalism, or other theories which accept the distinctive key element of that view. Certain writers place the idea of public justification at the heart of their work, whilst simultaneously rejecting Rawls's political liberalism.[384]

When the ideal of publicity obtains for a society, public justification is achieved. In the later philosophy, the ideas of public reason, and the liberal principle of legitimacy are

380 For an illustration of how we could conceive of principles of justice being supported by a society without the members of that society being moved by those principles itself, see Cohen (2008) pp. 127—129.
381 See *JF*, p. 185—186, *PL*, pp. 142—144
382 See fn389 below
383 Waldron (1993) pp. 57-58
384 One such writer is Gerald Gaus. See, for example, his (1996) pp. 3-5 for the endorsement of public justification, and (1996) pp. 131—136 and (2003) chapter 7, for the rejection of Rawls's political liberalism.

also introduced. Public reason, briefly put, is the body of knowledge, methods of inquiry, reasons and justifications from which are specified "the basic moral and political values that are to determine a constitutional democratic government's relation to its citizens and their relation to one another."[385] Public justification is hence to be achieved through the use of public reasoning.

Given the nature of liberal democracy as a political order, Rawls assumes that public justification is needed in order for the institutions and constitution of society to be legitimate. A political order, in anything other than an anarchist society, is an expression of state power. State power, in an ideal liberal democracy, must ultimately only be wielded by citizens as a collective body. The apparatus of the state, which constitutes a huge resource of technological know-how and institutional machinery, must be used only in ways which can be given public justification. Hence, the ideal liberal state as an entity is not conceived as anything over and above the citizens of the state and the state apparatus taken together in conjunction. But state power – liberal or not – is always coercive power, backed by sanctions.[386] Intuitively, this is enough to urge the need for legitimacy. But beyond even this, the political structure and basic institutions of a society impact on the character and aims of those who develop under them in profound and deep ways. Such a great impact also calls for justification. The ways in which the social order influences our upbringing must be capable of some appropriate kind of reflective endorsement by each member of society when they reach maturity.[387]

The liberal principle of legitimacy, as Rawls formulates 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."[388] Public reason is the common human reason spoken of here, or at least a major section of it. Any liberal view which endorses the idea of public justification – which aims to bring about a liberal regime governed by the right reasons, meaning public reasons[389] – has need of a similar principle of legitimacy. I'll call these *public justification*

---

385 *PL*, pp. 441—442, *CP*, p. 574, *LP*, p. 132

386 *PL*, pp. 68, 136, 216—217, *CP*, p. 482

387 *PL*, pp. 68, 269—271, *TJ*, p. 7/6—7. See also *TJ*, pp. 514—519/451—456, on autonomy, which at least in part is conceived as a kind of reflective endorsement of how our upbringing influenced the development of our character.

388 *PL*, p. 137

389 Not all liberalisms endorse the ideal of public justification, or place it at the centre of the idea of liberalism. These include certain Hobbesian liberalisms, certain value—pluralist liberalisms, and certain perfectionist liberalisms. In this taxonomy I have followed Quong (2011) pp. 12—21. I do not say that all varieties of these liberalisms must reject or downplay the importance of public justification, but at least some examples of each do so. See in addition fn108 below.

liberalisms.[390]

So the well-ordered society of Justice as Fairness must be governed by the right reasons: public reasons. But the society must also be stable, and stable for those reasons. The well-ordered society must not only possess a public conception of justice. It must also persist, given favourable conditions. And it must persist, given favourable conditions, due to its public conception.

The idea of stability spoken of here is essentially the same as that outlined in subsection 3.2 and elsewhere. I summarise it again here. It can be boiled down to two elements. The members of the well-ordered society must be able to realise a moral sensibility the content of which is given by the public conception of justice. As we have seen, this sensibility consists in the two moral powers. Furthermore, this sensibility must be compatible with human nature more generally. This means that the psychological strength of the motivations incorporated into the moral powers, and the strength of the other motivations which are congruent with those powers, must win out, at least in favourable conditions, against the strength of whatever further motivations human beings may be capable of developing under the institutions of the well-ordered society. Realising such a moral sensibility means that the members of the society grasp and are moved by the right reasons. The greater strength of these motivations, in comparison to opposing, unjust motivations, makes the society stable.

The account of the correct public conception, and the correct account of a stable well-ordered society, are interdependent. It should be remembered from chapter 2 that a well-ordered society must be sufficiently stable in order for a conception of justice to be justified, all-things-considered. To recap, a proposed conception of justice can fail in the argument from the original position in three ways. First, its principles can directly fail to meet the fundamental interests of the representees of the parties in the original position. But even if this test is met, a proposed conception of justice can fail if, second, it fails to be associated with a suitable moral psychology capable of (1) being realised by human beings, and (2) winning out against the special psychologies. If it is unable to meet one or both of these requirements, it would hence be futile. Third, and finally, it may be that all these requirements may be met, and so the conception of justice will be likely enough to be stable, given favourable circumstances. But it may be that the conception is comparatively less stable than some other conception which meets all the same criteria. It may come with

390 This moniker follows Quong (2011) p. 17, though on his usage public justification liberalism is a variety of political liberalism. I have restricted the term “political liberalism” to Rawls's view, or those positions which share the distinctive basic commitments of Rawls's political liberalism. These are outlined in subsection 14.1. I have not used Gaus's “justificatory liberalism” as I believe this is better reserved for Gaus's own view. See, for example, Gaus (1996), (2003) chapter 8.

a higher risk of being unstable, as was considered in chapter 2. The alternative conception will then win out in the parties' overall judgement, and the conception will have failed the test of arbitration.

If a given set of principles of justice fails the futility test,[391] then even if a society manages to come about which is governed by the corresponding public conception of justice, we can assume that many members of that society will eventually be moved to act unjustly. The ones who possess sufficient power will succeed. In so far as these powerful individuals use their power to alter the basic structure of society, the institutions of the society will become unjust. They will no longer be an expression of the collectively exercised power of the people, constrained by the requirements of public justification. The institutions will be an expression of the will of the most powerful factions in society. This amounts to the well-ordered society being unstable, for by definition a well-ordered society is governed and its power employed in accordance with its public conception of justice. Hence that conception of justice is unable to be sustained, even under ideal conditions.

By the requirements of the argument from the original position, this means that that conception of justice is unjustified, all-things-considered. But this means that the principles of that conception *cannot* provide the right public reasons by which a well-ordered society is to be governed. Even though we can imagine that this conception is capable of being *realised* as public conception, it is incapable of being *sustained*. This means it is unstable. But therefore, such a society was never governed by the right public reasons in the first place, as the right reasons must always be stable reasons over time.

In summary, stability for the right reasons obtains *only* for conceptions of justice which can be public conceptions shared between the members of the well-ordered society *and* which are stable over time. Reasons cannot be the right reasons without also being stable. And a conception of justice cannot be stable for the right reasons without its well-ordered society being governed by a public conception of justice. Hence, as was just stated, the account of the correct public conception of justice and the account of the stable well-ordered society are interdependent. It is important to get this dual-criterion on any conception of justice right in order to be clear on why the issue of stability for the right reasons leads to the revision of Justice as Fairness into a politically liberal theory.

Section 12: The Road to Political Liberalism

391 I leave aside here complications arising from considering the arbitration test.

In *Political Liberalism*, Rawls indicates how the commitment to stability for the right reasons led to the revisions of Justice as Fairness, and his endorsement of the idea of political liberalism. But his exact reasons are not fully or properly elaborated. To begin, I first introduce the ideas and modifications Rawls introduced into his theory in order to turn it into a politically liberal theory (subsection 12.1). Given this account, I then assess Rawls's statements as to why commitment to stability for the right reasons led him to these revisions and additions to his theory (subsection 12.2). I argue that these statements are in part misleading. But, whatever his exact reasons, Rawls's commitment to public justification and to what he calls reasonable pluralism is enough to lead to these revisions in any case.

12.1 Basic features of political liberalism

To begin to outline the idea of political liberalism, I first note that Justice as Fairness (and the internal problems which Rawls perceived in it which led to its revision), and the idea of political liberalism more generally, are distinct. It was reflecting on problems in Justice as Fairness which led Rawls to political liberalism.[392] But the general idea can be reached by different routes, and formulated in subtly different ways.[393]

There are several ways we might choose to introduce the idea of political liberalism. I start from an assumption Rawls makes about the institutions of liberal democracy. The institutions definitive of the ideal of liberal democracy – the fundamental liberal rights and liberties, and the arrangement of social institutions such as to guarantee every member of the society the ability to make effective use of those liberties[394] – inevitably give rise to a diversity of different comprehensive conceptions of the good, comprehensive religious and philosophical doctrines, and comprehensive ways of life.[395] Liberal democracy is marked by permanent pluralism.

A comprehensive conception or doctrine "includes conceptions of what is of value in human life, as well as ideals of personal virtue and character, that are to inform much of our nonpolitical life (in the limit our life as a whole)." Furthermore "by definition, for a conception to be even partially comprehensive, it must extend beyond the political and

392 *PL*, p. xv—xvi
393 Rawls cites (*PL*, p. 374 fn1), Larmore (1990) and Shklar (1989) as other writers who can be counted as political liberals, but who reached the idea by different routes. Neither are Rawlsians, particularly Shklar.
394 *PL*, pp. xlvi, 6
395 *PL*, pp. 3—4, 36, 39, 63—64

include non-political values and virtues"[396] (what a non-comprehensive, i.e. political, conception amounts to is outlined below). Comprehensive conceptions can be both liberal and non-liberal in their values and outlooks, and can be either secular or religious.[397] There have always been many comprehensive doctrines vying against one another.[398] In a liberal democracy, the diversity of opinion can be expected to be greater than under other regimes, due to the guarantee of the basic liberties, including freedom of thought, conscience, speech and association.[399]

Given that liberal institutions foster and promote such diversity, the question is whether this diversity is compatible with the long term sustainability of liberal democracy. Are the institutions of liberal democracy self-defeating? Do they give rise to too much diversity, such as to make it likely that too many comprehensive doctrines will arise which reject the basic liberties and ideals of liberal democracy, and do what they can to undermine them?[400]

In order for this outcome to be avoidable, it must be the case that there are not merely comprehensive doctrines, but reasonable and unreasonable comprehensive doctrines.[401] A wide enough selection of the doctrines which arise under the institutions of an ideal liberal democracy must be reasonable, or else be capable of reforming themselves to be reasonable if treated reasonably by already reasonable doctrines. Such doctrines are held to reform themselves by the lights of their own traditions of reasoning and reflection – not by political pressure or coercion.[402] Reasonable comprehensive doctrines are defined as those which are willing to wholeheartedly support the institutions of liberal democracy,[403] providing that they have good assurance that enough of their fellow citizens are willing to support those institutions as well.[404] Reasonable doctrines act reasonably in favourable circumstances. In unfavourable circumstances, in which they are threatened by many unreasonable doctrines, they still wish that they could behave reasonably. To the extent that it is possible, they try to act so as to bring about the liberal institutions which accord

396 *PL*, p. 175. See also p. 13
397 *PL*, p. xxxviii—xxxix
398 See *PL*, p. 37
399 *PL*, pp. 36, 63—64
400 See Rawls's statement of the "fundamental question" of political liberalism, *PL*, pp. xvii, xxxvii, 3—4. See also *CP*, pp. 620—622
401 *PL*, pp. xxxvii—xxxix, lviii—lx,
402 *PL*, pp. 65—66. On the idea that conceptions of the good must come to be reasonable by reflection using their own traditions of reason, see, for example, pp. 36—37, 169, 386—387. I note that p. 160 fn25 is misleading in this respect when says that a politically liberal conception of justice "shapes comprehensive views to cohere with it." This makes it sound like becoming reasonable is simply a psychological or sociological process which happens over time. Rather, the comprehensive doctrines must reason their own way to being reasonable.
403 *PL*, pp. 38—39, 58—61
404 *PL*, p. 49, 54

with their commitment to liberal democracy in the future.[405] As should be stressed, many comprehensive conceptions – again both liberal and non-liberal – are assumed to be reasonable.[406]

The reasonableness of reasonable persons, groups, and doctrines was partially described in subsection 8.2. We can expand on that characterisation here. Earlier I stressed that reasonableness consists in the willingness to propose fair terms of political organisation and cooperation. A second aspect of reasonableness consists in a willingness to recognise the burdens of judgement.[407] These burdens, as an aspect of the circumstances of justice, were mentioned previously in section 7. But they have yet to be elaborated fully.

The burdens of judgement are "the sources, or causes, of disagreement between reasonable persons."[408] Though reasonable persons are motivated to be conscientious and to come to fair agreement with each other, nevertheless for various reasons disagreement can be expected to persist between them. Between reasonable persons, such disagreement is reasonable disagreement. What is to be avoided is the idea that *all* disagreement stems from "most people [holding] views that advance their own more narrow interests" or from the fact that "people are often irrational and not very bright."[409] These represent sources of unreasonable disagreement, not reasonable disagreement. To grasp what is involved in reasonable disagreement, we must bear in mind "the different kinds of judgements"[410] that reasonable and rational persons are required to make. Rational persons must balance their various ends and assess their overall plan of life. Reasonable persons recognise they must assess "the strength of peoples' claims, not only against [their own] claims, but against one another, or on [their] common practices and institutions"[411] and also assess their "use of [their] theoretical (and not [only their] moral and practical) powers."[412]

But even when people recognise these complexities, and are conscientiously trying to come to reasonable agreement, there are a host of reasons as to why we might expect our judgements to nevertheless diverge. Jonathan Quong summarises them clearly.

> (*a*) empirical and scientific evidence may be complex and conflicting, (*b*) people may disagree about the relative weight that different considerations

405 See, for example, *PL*, p. 54, where Rawls says that the requirements of reasonableness, like Hobbes's laws of nature, bind "*in foro interno*". See Hobbes (1651) p. 110
406 E.g. *PL*, pp. 170
407 See *PL*, pp. 54—58, for this first feature, and pp49—50, for the second
408 *PL*, pp. 55—56
409 *PL*, p. 55
410 *PL*, p. 56
411 *PL*
412 *PL*

> should carry, (*c*) all conceptions are to some extent inherently vague and subject to hard cases, (*d*) the way we assess moral and political values is inevitably shaped to some degree by our total life experience, (*e*) there are often different kinds of normative considerations on both sides of a question which fully rational people may not agree how to place, and (*f*) social institutions are limited in the number of values they can incorporate, which will sometimes necessitate difficult or even tragic choices.[413]

The burdens of judgement are hence the explanation for the permanent *reasonable* pluralism found in even an ideal liberal democracy.[414]

Though they are divided in their comprehensive conceptions and doctrines, reasonable persons desire to live together with other reasonable persons on terms that are reasonable and fair. As was outlined in subsection 9.2, reasonable persons recognise each others' individual sovereignty, and the need for cooperative endeavour to be fair, for its own sake. To organise their joint affairs, however, they need to develop liberal conceptions of justice. They need articulated principles and reasons of sufficient determinacy to come to agreement on their political arrangements, or at least to narrow disagreement.[415] Such conceptions of justice are political conceptions of justice, in contrast to comprehensive doctrines and conceptions. Political conceptions are moral conceptions worked out to apply only to a specific subject – the political institutions of a liberal democracy, and to articulate their political ideals, including justice and legitimacy.[416] One difference between a political conception of justice and a comprehensive doctrine is hence down to scope. The former applies only to the political relations between citizens, whilst comprehensive doctrines may apply beyond this, to the limit of the whole of the universe and reality.[417]

There are many such liberal political conceptions of justice. Justice as Fairness is just one of them. They are bound together as a class by three common features. They assign certain basic rights and liberties to citizens. They assign a special priority to those rights and liberties "especially with respect to the good and perfectionist values." And they

---

413 Quong (2011) p. 37, citing *PL*, p. 55—57.

414 *PL*, p. 36—37. Several authors have questioned Rawls's account of the burdens of judgement, and the use he puts it to. See, for example, Wenar (2003) pp. 64—69. It is inessential for me to address such claims here. Quong (2011) pp. 245—246 thinks that Rawlsian-style political liberalism can be defended without the burdens of judgement, provided we maintain that "normal human reasoning ... under liberal conditions produces permanent disagreement," which is the position I assume here.

415 See *PL*, pp. 9—10, 156, 161, 223—227

416 *PL*, pp. xxxvi, xliii, 11—12. Note that justice and legitimacy are distinguished, as a legitimate government may still enact unjust laws (pp. 427—428)

417 *PL*, pp. 12—13

attempt to ensure that all citizens receive adequate all-purpose means to their liberties.[418] I shall refer to these three features as the three basic features of liberal conceptions. It is subscription to these three features as necessary requirements on their political arrangements which picks out comprehensive doctrines as reasonable. If a doctrine does not wholeheartedly endorse these three requirements at some level, whatever else it may hold, it is not reasonable.

Political conceptions are further distinguished from comprehensive conceptions in that they are developed purely within public reason, and using public reasoning.[419] Being developed within the limits of public reason, the political conceptions, *as a class*, are able to be the focus of an overlapping consensus of reasonable comprehensive doctrines. To repeat: an overlapping consensus need not be focused on a single liberal political conception, but may be focused on several.[420] The idea of an overlapping consensus is distinguished from the related idea of a *modus vivendi* in the following way.

A *modus vivendi* arises when "a plurality of conflicting comprehensive doctrines" are faced with certain historical circumstances that have "turned out [such] that for the time being at least, the balance of forces keeps all sides supporting the current arrangements which happen to be just to each of them."[421] There is hence some kind of consensus around shared political institutions. But each comprehensive doctrine in the society supports the political settlement merely instrumentally. It represents for them the best they can hope for given the roughly equal share of power between themselves and the other comprehensive doctrines. If the power balance shifts, no side will feel compunction in reshaping the society to accord with their comprehensive ideals. A *modus vivendi* hence occurs only

418 *PL*, pp. xlvi, 6, 223

419 See, for example, *PL*, pp. 8, 223—227. Rawls argues that political conceptions are in addition restricted to constitutional essentials and matters of "basic justice", e.g. *PL*, pp. 227—230. Not all political liberals follow him on this. See Quong (2011) chapter 9. I do not believe my basic arguments are effected whichever position is adopted, though there may be more influence when we consider how psychological theory interacts with the requirements of public reason (subsection 15.2).

420 *PL*, pp. 44 fn46, 149, 482—483, *CP*, p. 608—609, *LP*, pp. 172—173. Conceiving of the overlapping consensus as focused upon the group of political conceptions, is suggested by these passages. But this idea is not spelt out by Rawls himself. Quong (2011) chpt 6 represents a development and defence of the idea. I think he is probably on the right track, though I have not had the time to reflect on his position sufficiently to tell whether I would want to endorse it myself, or another similar to it. I have not incorporated this idea into the account of the stability argument in subsection 15.2 below. The way that Rawls presents that argument does not reflect this feature of the focus of an overlapping consensus, but instead restricts itself to talking of an overlapping consensus focused on Justice as Fairness. Given that my aim in that section is largely to indicate the continuity in Rawls's understanding of the stability argument between the later and earlier philosophies, incorporating the complexities of the overlapping consensus having a wider focus would be a distraction here. In addition, on Quong's own view, the overlapping consensus is no longer situated within the second part of the argument from the original position (Quong (2011) p. 186), but is rather presupposed by the fundamental ideas.

421 *PL*, pp. xl—xli. See also pp. 146—7. By "happen to be just to each of them" Rawls appears to mean that the arrangements satisfy at least some of the requirements of liberal political conceptions, not that the arrangements are just from the perspective of each comprehensive doctrine, for some of these may be unreasonable.

between unreasonable comprehensive doctrines, or between unreasonable and reasonable comprehensive doctrines.

The overlapping consensus of a well-ordered society, by contrast, "consists of all the reasonable opposing religious, philosophical, and moral doctrines likely to persist over generations and to gain a sizeable body of adherents."[422] In an overlapping consensus, "the acceptance of the political conception is not a compromise between those holding different [comprehensive] views."[423] Rather, the reasonable comprehensive doctrines affirm their preferred political conception for its own sake.

As we stressed in sections 8 and 9, in a well-ordered society members of the society value political justice and the political institutions of their society intrinsically.[424] Come the shift to political liberalism, the account of this becomes more complicated. First, each political conception is accepted by the reasonable members of the well-ordered society as giving *pro tanto* reasons for its own endorsement.[425] These *pro tanto* reasons are public reasons, as a political conception is, as has been said, purely developed from public reasoning. Reasonable comprehensive doctrines are then to determine for themselves how the *pro tanto* political reasons stemming from the political conception they endorse are to fit within the structure of their overall comprehensive view. It is imagined that there are many ways in which different comprehensive doctrines can do this, and the political conception itself gives no guidance as to how to it is to be conceived of as compatible with reasonable comprehensive doctrines.[426] How these added complications alter the stability argument will be reviewed in subsection 13.1 below.

I now move on to introduce what is often thought of as the most notorious aspect of Rawls's political liberalism: the idea that a liberal political conception is developed and presented as reasonable, and is not claimed to be true.[427] The first question to ask ourselves is: developed and presented by who? Individual political conceptions will be developed by particular individuals, such as Rawls, or they may be developed by politically active groups, or perhaps certain professions, such as the legal profession, or certain traditions, such as various religions, or even by political parties. A fully articulated and systematic political conception is the sort of thing we might expect to be developed by a philosopher or similar academic, but we might imagine a political conception which obtains only over a

422 *PL*, p. 15
423 *PL*, pp. 170—171
424 *PL*, pp. xxxvii—xxxviii makes it clear this this reaffirmed in the later philosophy.
425 *PL*, p. 386.
426 See *PL*, pp. 168—171, 386—387
427 See clear statements at *PL*. pp. xx, 94, 394—395. Comments by Raz (2003b) p. 396 serve to underline just how surprising we might find Rawls's position.

certain area of political consideration, such as the interpretation of a constitution, or on civil disobedience. The actual political conceptions endorsed by many individuals will most likely be a ramshackle collection of different political conceptions, ideas, reasons and principles – and most likely none the worse for it.[428] But really, all such exercises in reflection stem from a common source: the public reason of a liberal democracy. So when Rawls say that he does not put forward his *theory* as true, we should not focus on this. What we should understand is that Rawls thinks that the body of public reason in a liberal democracy should not be understood simply as true, in its own terms, but merely as reasonable.

What exactly are reasonableness and unreasonableness, in this context, such that they can be contrasted with truth and falsity? The True – i.e. the body of truths about the world – Rawls assumes, is unitary.[429] It is opposed to the False. But the False is not unitary. It is plural. There are many false things to think. But *both* the Reasonable and the Unreasonable are plural. There are many reasonable opinions, and only a selection of them could be true. The many reasonable beliefs oppose the many unreasonable beliefs. Again, only a selection of the unreasonable beliefs could ever be true. Truths and falsehoods, and reasonable and unreasonable beliefs, are contraries. Each one of this pair of contraries – true, false, reasonable, and unreasonable – is a different category, and obtains over a different range of beliefs or claims.

However, as Rawls understands these notions in this context, reasonableness is a more substantive notion than truth (though exactly what it means to say something is "more substantive", and to deny that it is simply formal, is, I recognise, a deep philosophical question). The sense of reasonableness Rawls means is that developed here and previously in subsection 8.2. From everything that has been said, we can say that the sense of "reasonable" meant is a specifically political sense of reasonable. This is required by the limitations of developing a non-comprehensive conception of justice and legitimacy suitable given the reasonable pluralism of ideal liberal democratic societies.

There can be assumed to be, similarly, a body of truths specifically related to politics and political philosophy. Political liberalism in no way aims to stop us from saying or thinking this. Assuming the unity of truth, there can only one truth about the way that a given society should be organised politically. But there can be many reasonable ways a liberal democracy can be organised. The reason why political reasonableness is, nevertheless, the more substantive notion than political truth is this: simply by saying that

428 For similar reflections, see *PL*, pp. 159—160
429 *PL*, p. 129

there is only one true way to organise a polity, we do not say whether this one true way is liberal, or democratic, or autocratic, or anarchic (i.e. to properly organise a state, we destroy it).[430] But all the reasonable ways to organise a society, given Rawls's usage, are liberal democratic ways to organise a society. All such ways are committed to the three basic features of a liberal regime. Hence political reasonableness and political liberalism are linked together for Rawls, and this gives his notion of reasonableness quite substantive content, even though he refrains from relating it to the notion of truth.[431]

Rawls believes that it is important that political conceptions are put forward as reasonable, and are not put forward as true, for a further reason. When put forward as true, a political conception can be in contradiction with a great many comprehensive doctrines, both reasonable and unreasonable (and, as it happens, with a great many political conceptions).[432] But it is unnecessary, for various reasons, for a liberal political conception to claim such a status – to claim it is true. For example, Rawls speculates that in proposing itself as true, a political conception might prevent itself from being the focus of a reasonable overlapping consensus. Hence, a liberal political conception should be content to put itself forward as reasonable.[433]

I shall briefly summarise the elements of political liberalism this subsection has introduced. Liberal democracies are marked by permanent pluralism in comprehensive moral doctrines. But this pluralism is not simply pluralism as such. It is reasonable pluralism, as much of the disagreement arises between conscientious reasonable persons.

---

430 See *PL*, p. 128, which clearly leaves this issue open.

431 See, for example, *PL*, p. 94. Comments on pp. 374—375 indicate that politics and political philosophy form a distinct subject matter, whether we are talking about political truth or political reasonableness. See also *LHPP*, pp. 3—5

432 See, for example, *PL*, pp. 126—127

433 There is much to be said about what putting forward a political conception as reasonable and not as true actually amounts to. I understand this to have both a metaethical aspect and a normative aspect. Any political conception will be accompanied by a metaethical story about the status of its principles of justice. For example, Rawls's liberal political conception is presented with a constructivist metaethic. But this metaethical story, whatever it is, needs to be publicly presented as reasonable, and not as true. I believe that we can best understand what it means to present the metaethical position underwriting a political conception as reasonable and not as true by understanding it as being public presented only as *prima facie*. Individuals then decide from the perspective of their own comprehensive doctrines whether they are going to accept this account as anything more than *prima facie*, or else abandon it in their non-public beliefs. The normative aspect of presenting a political conception as reasonable and not as true is that the normative authority of a political conception is only to be publicly asserted as *prima facie*. Whether the political conception has genuine normative authority, either *pro tanto* or all-things-considered, is again to be decided by citizens in the well-ordered society from the perspective of their comprehensive doctrines. To fully defend this interpretation of the idea of comprehensive doctrines being publicly presented as reasonable and not as true, and to critically assess it in light of how it relates to our interest in moral psychology in political liberalism, would be a lengthy task. For example, I have stated that the normative authority of political conceptions is to be presented as *prima facie*, but I have also (in the main text) stated it is presented to reasonable comprehensive doctrines as *pro tanto*. I have not explained here how these two claims are consistent. But I put it aside this and other issues for the course of this discussion. Hence, for the rest of the chapter I will simply refer to political conceptions of justice being presented as reasonable and not as true without further explaining what this means.

The reason for this is that there are burdens of judgement which effect even the debates and disputes of those who are wholeheartedly trying to be reasonable. Comprehensive moral doctrines can be both reasonable and unreasonable. The former endorse the basic institutions and ideals of a liberal democratic regime. The latter do not. Being divided in their comprehensive moral doctrines, but at the same time being subscribers to the ideal of liberal democracy, reasonable comprehensive doctrines – and politically reasonable persons – develop political conceptions of justice. Political conceptions are marked out by their scope. They are moral conceptions which apply only to issues regarding the political arrangements of liberal democracies. Political conceptions are also marked out in that they are drawn entirely from the public culture of liberal democracies, and hence are exercises in public reasoning. In a well-ordered society, political conceptions, taken as a group, are the focus of an overlapping consensus of the reasonable comprehensive doctrines. The political conceptions provide *pro tanto* reasons for their acceptance to reasonable comprehensive doctrines. The full account of how such political values and principles fit with a person's other values and principles is given by how their particular comprehensive doctrine fits the former values and principles in with the latter. Finally, political conceptions should ideally be understood as reasonable rather than true. There can be many political conceptions, and many comprehensive moral doctrines, that are reasonable. Assuming Truth is One, there can only ever be one body of truth regarding moral and political matters. But the fact of reasonable pluralism means that political conceptions proposed as true are not suitable as the focus of an overlapping consensus.

*12.2 Why was **Theory's** well-ordered society not stable?*

Having laid out the fundamentals of the idea of political liberalism, and earlier the notion of stability for the right reasons, I now investigate what reasons Rawls had, or may have had, for revising Justice as Fairness. I first present Rawls's account. I then argue that aspects of this account are misleading, in particular the idea that the alterations to Justice as Fairness were necessary in order to make its well-ordered society realistic. This is true, but the revisions were needed anyway, simply in virtue of the combination of the need for public justification, combined with the fact of reasonable pluralism.

Rawls believed Justice as Fairness failed to be stable for the right reasons, in its original comprehensive formulation. Hence why he held it needed to be reformed. Justice as Fairness failed because of the facts that political liberalisms more generally accept. It did not recognise the fact of reasonable pluralism which characterises modern liberal

democracies, even when they are conceived in their ideal form. But let's see how Rawls puts the matter himself.

Rawls writes that the problem arises due to the "unrealistic idea of a well-ordered society as it appears in *Theory*."[434] In this original formulation of the well-ordered society, all members of the society endorse Justice as Fairness and its two principles "on the basis of ... a comprehensive philosophical doctrine."[435] But as we have seen, "a modern democratic society is characterised by ... a pluralism of incompatible yet reasonable comprehensive doctrines" which, being reasonable, do not "reject the essentials of a democratic regime."[436] Justice as Fairness is, in its original formulation, a comprehensive liberal conception of justice. It also accepts those essentials. But this leads to the difficulty:

> However, since the principles of justice as fairness in *Theory* require a constitutional democratic regime, and since the fact of reasonable pluralism is the long-term outcome of a society's culture in the context of these free institutions, ... the argument in *Theory* relies on a premise the realisation of which its principles of justice rule out. This is the premise that in the well-ordered society of justice as fairness, citizens hold the same comprehensive doctrine, and this includes aspects of Kant's comprehensive liberalism, to which the principles of justice as fairness belong. But given the fact of reasonable pluralism, this comprehensive view is not held by citizens generally, any more than a religious doctrine, or some form of utilitarianism.[437]

The idea is that the society of Justice as Fairness, when conceived as based on a comprehensive conception as in *Theory*, cannot be realised. Its institutional structure gives rise to reasonable pluralism, and not the monism of Kantian comprehensive liberalism which Justice as Fairness in *Theory* is described here as helping itself to when arguing for the stability of its well-ordered society. To expect all members of the society to be, in some sense, comprehensive Kantians is "unrealistic,"[438] even in favourable conditions. But if this is unrealistic, then the argument in *Theory* is not completed. The stability argument as presented in that book does not by itself secure the stability of the institutional structure of

---

434 *TJ*, p. xvi. See also *CP*, p. 488—490 which speaks of the society of *Theory* as being "utopian", in an unrealistic, and not realistic way (on this latter distinction, see *JF*, pp. 4—5). See further subsection 13.1 below.
435 *TJ*, p. xvi
436 *PL*, p. xvi
437 *PL*, p. xl
438 *PL*, p. xvii

the society described. As was made clear in section 11 above, and in earlier chapters, such stability is required for full justification. Of course, such a society might be stable, in a certain sense, if those who accept the comprehensive version of Justice as Fairness impose conformity to its requirements on all other members of the society. But this would require the violation of basic liberal rights. The society would be stable for the wrong reasons. This is what Rawls asserts when he writes that "a society united on ... the reasonable liberalisms of Kant or Mill, would ... require the sanctions of state power [to be used in an oppressive, illegitimate manner] to remain so [united]."[439]

This introduces some of the reasons for Rawls's alteration of his views. But my first point is that simply saying that the well-ordered society of Justice as Fairness was unrealistic, or that it would require a repression which would amount to the contradiction of its very own principles, is potentially misleading. What may be suggested is that Rawls simply views the well-ordered society of Justice as Fairness as psychologically unrealistic, i.e. unrealisable or unstable in the light of human nature. The changes in Rawls theory would hence be forced by Rawls reassessing his earlier account of human nature.

However, what becomes clearer from examining Rawls's position as a whole is that the essential problem isn't just that we could not bring about the well-ordered society of *Theory*. It's that we should not even want to. What is important is not that the society of *Theory* is psychologically impossible, but that it would not be stable for the right reasons even if it were possible. As described in *Theory*, the well-ordered society is unjustified. It is either unjustified because its argument from stability is incomplete, or it is unjustified because elements of its argument from stability are in contradiction with one of the other elements of the theory – the priority of liberty. Assuming a close connection between the priority of liberty and the fundamental conception of the person (not a rash assumption, I feel), the argument from stability is also in contradiction with the assumptions of the conception of the person.

To see this, let's imagine, against the assumptions from the burdens of judgement, that the well-ordered society of Justice as Fairness in *Theory* did come about. The members of the society have a public conception of justice, which they follow and which sustains the society in perpetuity. Is it the case that this society will be stable for the right reasons?

Now imagine that the members of the well-ordered society consider the following (for them) hypothetical scenario: what if other reasonable comprehensive moral doctrines – whether liberal or non-liberal – were to arise in their society? Being reasonable

439 *PL*, p. 37

comprehensive doctrines, these other comprehensive doctrines will endorse some liberal political conception of justice, even if it is not Justice as Fairness. Given the uniform endorsement of the comprehensive version of Justice as Fairness, the members of this society had not previously considered this possibility. Hence, consideration of it was not included within the justification of the public conception of the well-ordered society known to each member of the society. This situation is perfectly possible by Rawls's stipulations. Justification via the original position need not be thought to be definitive, but is simply the most justified conclusion we can offer given the progress of our reflections so far.[440]

We might imagine two possible reactions to such hypothetical moral musings, and two corresponding extensions of the public conception of justice of the society:

a: If more reasonable comprehensive doctrines arose, it would be justifiable to suppress them using the coercive power of the state.

b: If more reasonable doctrines were to arise, it would not be permissible to suppress any of them, except if, in the case of certain ones, failing to suppress those ones would be certain, or highly likely, to lead to the liberal institutions of the society being undermined.[441]

These two reactions can be thought of as extensions of the existing public conception. The members of the society are formulating new beliefs, linked to their existing attitudes, on a scenario they have not previously considered. For the later Rawls, only societies which adopted **b** above would be classifiable as stable for the right reasons.

Societies which conform to **a**, though perhaps stable, are not stable for the right reasons. This is because their public conception of justice, which would contain **a**, could not be publicly justified to new reasonable comprehensive doctrines if any were to arise. Given **a**, new reasonable comprehensive doctrines would be suppressed. But how could this be justified to the advocates of those doctrines? After all, the new doctrines would endorse liberal democratic institutions. Being reasonable, they would be willing to offer justifications for this endorsement which would be public. This would require that those justifications be understandable by anyone else who endorses the ideal of liberal

440 See *TJ*, p. 52/45—46, 508—509/580—582, *JF*, p. 31

441 The question of whether and to what extent it is permissible to oppress *unreasonable* comprehensive doctrines is a complicated one. I do not take it up here. Rawls says something about it (see, for example, *TJ*, pp. 216—221/190—194 and 575—577/503—505), and the issue has recently received extended treatment by Quong (2011) chapter 10

democracy. The public culture and body of public reasoning of the existing liberal society, which in our scenario contains only advocates of Justice as Fairness in its comprehensive guise, would be expanded by the presence of these new doctrines. The existing inhabitants of the society could hardly claim that the new arrivals were appealing to reasons they simply could not recognise, given the shared commitment to liberal democratic institutions and basic liberal rights. Furthermore, it is important to note that these considerations apply even if the development of new comprehensive doctrines was truly counter-factual, and none actually did ever arise. For the members of the society who endorse **a** would be willing to suppress other doctrines if they did arise. But in holding this, they show that they are not genuinely committed to basic liberal rights, even though they never get a chance to demonstrate their illiberality. Though their society is stable, it is not governed by the right reasons – at least in one key respect.

Hence, whether or not we accept that the well-ordered society of *Theory* is unrealistic, its public conception is importantly undetermined, as it does not address the above possibility.[442] In so far as it fails to address this possibility, it is (to that extent, we might say) unjustified as a liberal conception of justice, providing we assume that a liberal conception of justice has to have something to say about reasonable pluralism.

Given all this, what can we say about Rawls's claim that the well-ordered society of *Theory* is unrealistic? The claim, I assume, is plausible. But it was a misstep for Rawls to put things primarily in this way. The problem with the well-ordered society of the comprehensive version of Justice as Fairness is not primarily that it is unrealistic given real-world circumstances (though this would be a problem). The more fundamental problem is that such a well-ordered society would be stable for the *wrong* reasons, even in favourable circumstances which allowed such a society to occur and thrive. Liberal democratic societies are marked, Rawls assumes, by reasonable pluralism. In order for there to be public justification of the institutions of such societies, and hence for them to be legitimate liberal states, there need to be specifically political conceptions of justice. This is the only way for there to be shared public reasons between all members of the society, as they are divided in their comprehensive moral doctrines, and the reasons associated with these. Hence, we need a revised stability argument: one which takes reasonable pluralism into account. A society united under one liberal comprehensive doctrine, if it could even exist, is simply *not* a liberal society, in political terms at least. It is perhaps better described as a liberal culture. Whether this liberal culture existed under a liberal or an illiberal state

442 *PL*, p. xv—xvi for a statement that *Theory* simply fails to take a view on this matter, i.e. does not explicitly come down on either **a** or **b**.

would depend upon whether its members would publicly endorse, on consideration, **a** or **b** above.

Given these considerations, I draw two conclusions. The first is that the fact of reasonable pluralism is important simply as a possibility, independently of whether it is realised or not. The fact of reasonable pluralism is of course assumed to obtain in our world. But if there were another world in which it did not obtain, reflection on the possibility of reasonable pluralism would still force the changes on Justice as Fairness which Rawls identifies. This follows simply from the requirement of public justification.[443]

The second conclusion is that problems with the psychological realisability of the original account of Justice as Fairness are not at the heart of the modifications in Rawls's later work. In both the earlier and the later philosophy, psychological realisability has a prominent place. What I want to suggest it is the same prominent place. It is not that a renewed concern with psychological stability, and Rawls's existing moral psychology, led to revisions to the idea of stability for the right reasons, and the introduction of the distinctive ideas of the later works. Rather, consideration of the idea of stability for the right reasons lead to these introductions directly. It is left an open question how much the moral psychology has to be altered in the light of this (which is not to say it will not be altered).

Contrary certain interpretations of his work,[444] Rawls's later philosophy does not show an increased concern with stability over justice. Rather, it notices certain overlooked ramifications of the endorsement of public justification and basic liberal rights, and attempts to develop what conclusions these may lead to. This does lead Rawls away from being concerned to develop the true theory of justice, but only due to self-imposed restrictions arising from what he understands to be the requirements of public justification. This is different from altering one's theory simply due to the practical limitations of the world *per se*. As I remarked in subsection 6.1, Rawls's conception of the person (and of the well-ordered society) represent non-negotiable normative foundations to his theory. If he were simply adjusting his theory to practical restrictions, this could not be true.

Hence, it was not problems with psychological realisability which primarily led to the revisions found in the later theory. Some of Rawls's own remarks are hence misleading in this regard. Approaches to explaining why Rawls revised his theory which focus on

---

443 Whether reasonable pluralism is to be understood exactly as Rawls understands it, or whether public justification requires the distinctively Rawlsian idea that political conceptions must be presented as reasonable and not as true, is left open by this conclusion.

444 For example, Klosko (1994). Interpretations which go against such positions include Krasnoff (1999) and Quong (2011) chapter 5. These latter interpretations are more in line with my own.

looking for weaknesses in the stability argument, such as Barry's,[445] are in a subtle way misguided. The stability argument is, in truth, the wrong aspect of Rawls's theory to start from in tracing the changes in his thought. The more basic idea of stability for the right reasons should be our point of departure, and it is this road which I have attempted to take in this section. In the next section, I shall begin to examine what changes may need to be made to the stability argument and Rawls's account of moral psychology given the alterations to his theory in his later period.

Before continuing however, I remark to the reader that I do not aim to defend political liberalism, or even public justification liberalism in this chapter. My aim is simply to clarify what stability for the right reasons commits Rawls to, in order have a firm footing from which to understand the place of moral psychology within Rawls's later theory. I hence leave aside here debates about whether political liberalism is or is not the only correct formulation of liberalism as a political philosophical theory.[446]

Section 13: Moral Psychology in Political Liberalism

In the previous section, I assessed why stability for the right reasons led to the revisions of the late period Rawls. In the previous subsection in particular, I argued that it is not particularly issues of psychological realism which led to Rawls's revisions of his philosophy, or at least which should have led Rawls to his revisions. Rather, it is the requirement of stability for the right reasons itself, when combined with the fact of reasonable pluralism. It is the need for public justification which leads to the need to distinguish between comprehensive and political conceptions. In this section, I try to make some in-roads into how much this may cause alterations in Rawls's moral psychology, as it appeared in *Theory* and other early works. My first claim is that the roles of moral psychology, and the shape of the stability argument, do not and need not change much at all. I then observe that the changes made in Rawls's later period do not in themselves require that the content of the psychology be changed at all. It may be that Rawls's moral psychology does need to be altered. But it is difficult to say whether this is so, without a much more substantial account of how the requirements of public reasoning limit the use of psychological data. This is not given by the basic ideas in Rawls's later work, and no more sufficiently substantial discussion of this issue is provided by him, as the discussion

445 Barry (1995b)

446 For representative exchanges regarding this issue, one for political liberalism, and one against, see, respectively, Quong (2011) and Wall (1998).

below shall illustrate.

I first briefly outline the stability argument as it now appears in the later works (subsection 13.1). This allows me to introduce the uses that Rawls put his moral psychology to in his politically liberal period, which have very minor differences with his comprehensively liberal times. It also allows me to observe that the essential roles of the moral psychology are unchanged. In the following subsection, 13.2, I then very briefly comment on how the limitations imposed by political liberalism may impact on the assessment of the moral psychology.

13.1 Rawls's use of moral psychology in his politically liberal theory

As I noted in the introduction to the thesis, the general shape of Rawls's moral psychology, and the use to which he puts it, does not significantly alter between *A Theory of Justice,* and *Political Liberalism* and beyond. In this section, I outline how the roles of moral psychology are essentially the same in Rawls's later theory as in his earlier.

In *Political Liberalism,* Rawls refers to his continued reliance on the moral psychology.[447] The moral psychology is again referenced, summarised this time in slightly more expanded detail, in *Justice as Fairness: A Restatement*.[448] In this second discussion, Rawls states that he "would not change ... substantially" the earlier account.[449] It should be noted, however, that he only makes reference to §§70-72 and §§75-76 of *Theory*. The sections missing are those that discuss the moral emotions, the moral sentiments, and the natural sentiments. I am unsure whether this omission has any deeper significance. Given the brief account of the moral psychology Rawls presents here, it may be that he thought these sections unnecessary to cite. Alternatively, he may have believed there were more fundamental problems with his views on these matters. I do not here investigate this matter, though it is of obvious significance regarding debates surrounding what sort of role both the early and late Rawls saw (and should have seen) the emotions having in politics.[450]

I have earlier noted in subsection 3.1 that the content of the moral psychology is broader than just the account of the sense of justice. It is clear that these further elements are also included within the moral psychology in Rawls's later views. The concept of the person is the same there, and the account of the ability to revise and develop a conception

447 *PL*, p. 143 fn9. Confusingly, this psychology is only summarised at pp. 86 and 163. See also *CP*, p. 445

448 *JF*, pp. 195—197, esp fn17

449 *JF*, pp. 195—197

450 On this debate, see, for example, Solomon (1995), Nussbaum (2003) pp. 489—499, Held (2006) pp. 83—84, Krause (2008), pp. 28—37

of the good is largely the same.[451] Indeed, I made use of this material in chapter 4.

In his later works, Rawls calls his moral psychology "a reasonable moral psychology." Regarding this, he writes that "this name is appropriate since the idea of reciprocity appears both as a principle giving its content and as a disposition to answer in kind."[452] I think the "principle" referred to here is most likely a normative principle. The "disposition" by contrast corresponds to the psychological principles of reciprocity: see Appendix II.

We can reconstruct from *Political Liberalism* and *Justice as Fairness: A Restatement* a subtly new account of the stability argument. It will be seen to employ moral psychology once again in its roles of demonstrating the realisability and stability of the well-ordered society, and also playing the justifying roles of avoiding futility and arbitration.

The "question of stability"[453] is answered in the following stages. First, the moral psychology of the sense of justice is again presented in order to demonstrate that, living under the institutions of Justice as Fairness, citizens acquire a sense of justice and the corresponding motivation.[454]

It can then be argued, as discussed in chapter 2, that the sense of justice generated by Justice as Fairness wins out against competing moral psychologies, developed this time from rival *political* conceptions. As has been made clear, political liberalism rejects the idea that a straight comparison between the strengths of the sense of justice associated with different comprehensive moral conceptions can serve in an argument for the stability (for the right reasons) of a liberal democracy, given the fact of reasonable pluralism. As previously noted in subsection 4.5, however, Rawls does not appear to stress such comparisons in the various restatements of the argument from stability in his later works. In addition, he cites the "relative stability" section of *Theory* as one of those which indicates that it assumes an unrealistic, monistic conception of well-ordered societies and their comprehensive doctrines.[455] It may be that he meant to indicate that assessing the comparative stability of different conceptions should be dropped from the argument in the original position, and hence also the role of arbitration. Yet the passage just referenced does not say this, but simply observes that this is one place in *Theory* in which Justice as Fairness is assumed to be a partially comprehensive doctrine, in contradiction to Rawls's later formulation of his theory. This does not rule out the comparison of different political

451 See, for example, *PL*, pp. 81—86, 176—178
452 *PL*, pp. 195—196. The name is introduced at *CP*, p. 445
453 *PL*, p. 140
454 See *PL*, pp. 140—143, *JF*, pp. 195—197
455 *JF*, pp. 186—187. See also *CP*, p. 489

conceptions along these lines.[456] And when we consider it, why should the parties in the original position deprive themselves of such potentially relevant considerations? I do not think it is obvious that Rawls needs to drop relative stability comparisons from the argument from the original position entirely, provided they are understood to be comparisons between the psychologies associated with rival political conceptions.[457] But I leave it open that he may have decided not to stress this element of the stability argument in later work.

Next, the special psychologies are again considered, and it is argued that, in the well-ordered society of Justice as Fairness, these psychologies will not be so powerful as to overwhelm our motivation to act justly and civilly as citizens.[458] Note that the reason that Rawls is able to include the discussion of the special psychologies here, rather than in the discussion of the overlapping consensus, say, is that he ties them closely to a variety of conflict which it is not political liberalism's job to specifically address. These are conflicts from "citizen's status, class position, and occupation, or from their ethnicity, gender and race."[459] It is a deeply complicated issue as to how far disputes between comprehensive doctrines can be disentangled from these further disputes, at least in our non-ideal circumstances. It may be that they cannot, and hence perhaps that political liberalism takes the wrong approach to these problems.[460] It is an important issue as to how successfully Justice as Fairness is able to deal with these issues, and if it is not successful, what this means for Justice as Fairness, and liberal theory generally. But I do not address this here.

To return to our current topic, arguments regarding the congruence of the Right and the Good are now reprised.[461] These sections are the ones most frequently cited as responsible for Rawls's revisions of his work in *Political Liberalism*.[462] I have reservations about the way such claims are usually put. But there is undoubtedly something correct about them – Rawls himself frequently calls attention to the difference between political and comprehensive conceptions of the good of liberal society.[463] But I will not air my reservations here, as noted in section 12 above. But it is clear that Rawls puts forward several arguments for the good of political justice, at least when viewed from the

456 As suggested by the reference to the relevant section at *JF*, p. 196 fn17
457 See, for example, places in which the account of the original position argument in *Justice as Fairness: A Restatement* may be referring to comparative stability considerations, e.g. pp. 124—130. I only say "may": recalling the various complexities chapter 2 highlighted about the interpretation of this material.
458 *JF*, pp. 184—185
459 *JF*, p. lviii
460 For an example of such a critique, see Okin (1994)
461 See *PL*, pp. 140—141 fn7, 142—143 incl. fn9
462 See, for example, Barry (1995b) pp. 885—890, Freeman (2003) pp. 303—308, (2007b) pp. 167—172, 178—186, Quong (2011) pp. 163—164.
463 See, for example, *PL*, pp. 97—99, 455—456 *CP*, p. 586 *LP*, p. 146

perspective of Justice as Fairness as a political conception.[464]

It is only after all these elements of the stability argument from *A Theory of Justice* have been reprised, now as elements of a political conception of justice, that the argument for the possibility for an overlapping consensus is taken up. This argument uses the same moral psychology of the sense of justice (and reasonableness) which has been used throughout the stability argument.[465] If it can be shown that an overlapping consensus is a reasonable possibility, given the wider social situation of the well-ordered society, then Justice as Fairness will be a stable political conception, and hence be justified all-things-considered.

It is important to note here is that the argument does not proceed by arguing that, given its moral psychology, Justice as Fairness *will* develop an overlapping consensus around itself. Rawls clarifies this in the introduction to the paperback edition of *Political Liberalism*. That book

> makes no attempt to prove, or to show, that [an overlapping] consensus would eventually form around a reasonable political conception of justice. The most it does is to present a ... liberal political conception that does not oppose comprehensive doctrines on their own ground and does not preclude the possibility of an overlapping consensus for the right reasons. *PL* does note certain historical events and processes that seem to have led to [a more limited] consensus, and others that may take place, but observing these commonsense facts of political sociology does not constitute proof.[466]

The idea of an overlapping consensus does not complete the stability argument by arguing that the institutions of the well-ordered society will lead to such a consensus, given the moral psychology of Justice as Fairness. Rather, it completes the stability argument by showing that it is not a foregone conclusion that the public acceptance of a liberal conception of justice is incompatible with the reasonable pluralism which basic liberal institutions inevitably give rise to.

This limitation on the stability argument is not imposed simply because it would be impossible to determine whether, psychologically, the institutions of Justice as Fairness give rise to an overlapping consensus. Rather, making such an argument would require us

464 See *JF*, pp. 198—202, *PL*, pp. 201—206, *CP*, pp. 465—470
465 *PL*, pp. 86 fn34, 141
466 *PL*, pp. xlv—xlvi. The same idea is much more subtly suggested by phrasing at *JF*, pp. 190—191

to go outside the limitations of public reason.[467] There may be several reasons for this. I give one here. To make such an argument would require a quite extensive social and political psychology, which would give a general theory regarding how liberal political institutions and a range of comprehensive doctrines with certain features generally interact. Such a psycho-social theory would no doubt be beyond the limits of public reason. Such a theory would not describe how comprehensive doctrines come to endorse the political conception in their own terms, but by reference to the psychological theory. But obviously, the advocates of such comprehensive doctrines couldn't accept the theory's explanation for their behaviour without abandoning their own views.[468]

In *Theory*, the moral psychology is restricted to demonstrating how, in a well-ordered society, individuals would acquire a sense of justice over the normal course of development. In *Political Liberalism, Justice as Fairness:A Restatement*, and the preceding article "The Idea of an Overlapping Consensus", the psychology is also used as part of an argument for how a society initially characterised by a *modus vivendi* between its different comprehensive doctrines may, given favourable conditions, develop into a well-ordered society.[469] This might surprise us. Why doesn't Rawls simply illustrate the possibility of various comprehensive doctrines within an well-ordered society endorsing the political institutions of that society for their own sake?

I speculate that this historical just-so story may be presented in order to help better justify political liberalism to ourselves. In our historical situation, Rawls believes the United States, and certain other democracies, to have achieved what he calls a constitutional consensus. But he perhaps believes that by telling a certain psychological story of how we might have come to such a consensus, and how we might, by the same kinds of social changes, come to an overlapping consensus, is enough to show that an overlapping consensus is not an incoherent or obviously outlandish idea, and hence is "not [unrealistically] utopian."[470] This allows Justice as Fairness, and the idea of political liberalism more generally, to be of practical relevance.

---

467 *PL*, p. 387 suggested this interpretation to me, though it is talking about justification, and not directly about stability.

468 It may be worried that Rawls's own moral psychology may face similar problems. Against this, there are two considerations. The first is that the moral psychology is required to be framed only using publicly accessible psychological claims (see subsection 13.2 below). The second is that it is to be presented as reasonable and not as true. Individuals are hence free to come up with their own explanations as to why they have been psychologically moved to endorse one of the liberal political conceptions, or even why everyone else has as well. I believe there may remain problems with Rawls's own presentation in this regard. In particular, Rawls sometimes talks as if the account of the overlapping consensus shows how the society of Justice as Fairness 'adapts' comprehensive doctrines it itself, as noted above in 12.1 (see also *PL* p. 219). I am unsure whether this kind of phrasing is compatible with the limitations of public reasoning.

469 Rawls's initial account occurs in *CP*, pp. 440—446 and . This is expanded in *PL*, pp. 158—168. See also *JF*, pp. 192—195

470 See *JF*, p. 192. See also *PL*, p. 158

With the presentation of the possibility of an overlapping consensus, the stability argument in the original position is now, once again, complete. The first, second, third and fourth roles of moral psychology are all found within this argument. I will close this subsection by noting that the fifth and sixth roles of moral psychology are still present in the later philosophy as well. Our capacity for the moral powers is still taken to be the criterion for being owed justice, as the next chapter will examine at length. Moral psychology may possibly still be partially constitutive of Justice as Fairness as before – though of course it can now only be asserted to be a constitutive element of a political conception.

13.2 Psychology and Public Justification

As noted in subsection 4.5, the parties in the original position in *Theory* were conceived to have access to all social scientific theory. But they were also conceived to have a preference for conceptions of justice which were supported by scientific and social-scientific theories the basic content of which was capable of being publicly accessible. Come political liberalism, however, this preference was converted into a requirement. The parties now have access only to those aspects of scientific theory which are publicly available and uncontroversial. How much might we expect this restriction to alter Rawls's moral psychology?

Judging by his own assessment, not much. As we have seen, Rawls indicates that the essentials, if not most, of his psychology are to be carried over from Justice as Fairness's comprehensive formulation. I am myself unsure to what extent his position can be defended. But I have these reservations not because I am sceptical about Rawls's claim, but because thinking about what restrictions public justification puts on the use of empirical psychological data is yet another substantial topic.[471] I have not, I feel, reflected sufficiently on this matter to come to any definite conclusions. As I said at the outset, my focus is on the roles of moral psychology within Rawls's theory, and not on the content of the psychology he makes use of, or how it might be assessed. But I here note why it seems that those who would defend Rawls cannot avoid this issue.

The parties only recognise psychological data which can be framed in terms of public reason. This will undoubtedly fall short of the claims of empirical psychological

471 As Rawls recognises at *PL*, p. 252

science,[472] or even the moral psychological claims which are developed to accompany comprehensive moral theories (remembering the distinctions made in subsection 5.2). Rawls glosses framing psychological results within public reason by saying that psychological results cannot be controversial. But this does not seem to specify things sufficiently precisely. Presumably it is possible to unreasonably discount some psychological data. It may be that certain psychological facts are controversial, to us, but only because they are being unreasonably discounted.[473] In which case, the parties in the original position would recognise them. How might we come to recognise which facts the parties would recognise and which they would not? Rawls tells us that the parties also agree to principles of public inquiry as well as justice, and that we can reflect on the original position to see what these may be.[474] But he does not elaborate much.

Furthermore, it does not seem to be the case that, if psychological debates turn out to be rampantly and pervasively controversial, then the parties in the original position could simply discount psychological facts. The finality condition, and the demands of stability, tell against this. Rawlsian-style contractualism must be able to recourse to the facts of human nature to some significant extent, if it is to succeed at all.

I will not consider this matter further. To conclude this chapter: I hold that the changes introduced by Rawls to Justice as Fairness in his later period do not, in themselves, obviously and directly mandate the alteration of substantial elements of his moral psychology. Whether this psychology would need to be altered would depend on the restrictions imposed by the requirements of public justification, which, it should be remembered, were already present in the earlier philosophy in a limited way. It would be hasty to assume that these alterations would be wide ranging. It would also be hasty to assume that they would not be. But the same can be said of the psychology presented in the earlier philosophy, if we decide to reject Rawls's later modifications of his theory. The ideas introduced to Justice as Fairness in the politically liberal period, considered in themselves, obviously place more restrictions on his psychology than were already present. But how extensive are these new restrictions, and how much do they add to the restrictions which were already in place in the earlier philosophy? The question simply cannot be answered, without a clear view of the admissibility of psychological data into the public forum.

472 Public reason, and scientific reason, are specified to be different at *PL*, p. 221

473 A similar theme is prominent in Gerald Gaus's work. See Gaus (1996), (2011). To what extent the recognition of this issue could be combined with a more Rawlsian approach is as yet under-explored.

474 See *PL*, pp. 223—227, *JF*, pp. 91—92

Chapter 6: Moral Psychology and The Scope of Justice

In this final chapter, I look at the scope of justice within a well-ordered society — to whom claims of justice are attributed, and from whom the requirements of justice are expected. Unlike previous chapters, this chapter is much more substantively critical, rather than exegetical. As noted in subsection 3.4, Rawls's account of the scope of justice alters throughout his career. I believe that his earliest account is the most defensible, though to defend it Rawls's later theory needs to be modified. On the earliest account, a sense of justice, and a capacity to develop a conception of the good, is necessary and sufficient for an individual to be included within the scope of the rights and responsibilities of justice. I defend this position, in part, by reference to aspects of Rawls's moral psychology. Section 14 introduces the discussion and sets out the various possible positions: the content of sections 15 through 17 are listed at its end. Section 18 then remarks on whether we should accept Rawls's account of the scope of justice, even in the modified form I have presented it. Though I believe that the modified position allows for a contractarian account of justice which is more plausible than has sometimes been thought, I have unavoidable reservations about it given the overall structure of Rawls's theory, which I shall outline.

Section 14: Turning back the clock on the scope of justice

I aim to give the best defence I can to the view that the capacity for the moral powers is necessary and sufficient to be owed justice. These powers, it will be remembered, are the sense of justice, and the capacity to develop a conception of the good (subsection 3.4). I believe that this is the most plausible view available to Rawls, given both what I view as the most essential aspects of his theory, and also my own moral and philosophical judgements. But though I give the best defence I can, my affirmation of this position will be found to be, at best, half-hearted. This is because I believe that serious problems yet remain, and are most likely to remain, as section 18 will discuss. This should be remembered throughout the coming discussion. Though I suggest alterations to Rawls's position, I work for the most part within a Rawlsian framework. It is only within section 18 that I then present my general misgivings about a Rawlsian position on these matters.

The Rawlsian position I defend is the one Rawls appears to have held in his earliest article to include a discussion of the scope of justice – "The Sense of Justice". Things changed after that, however. Defending this earliest position means that elements of the later formulations of the theory will need to be altered, or dropped altogether. These

revisions are not entirely external. Even subsequent to "The Sense of Justice", certain elements of Rawls's view make the most sense when combined with the position that the moral powers are necessary and sufficient to be owed justice. Most importantly for us, these elements include core elements of his moral psychology. My eventual position, then, is revisionary. At times it does follow my own moral assessment more than Rawls's. But I believe that *some* of Rawls's commitments are simply morally indefensible, so there is no choice in this matter. However, I attempt whenever possible to argue that Rawls's himself, given some of his *other* commitments, should have supported the revisions I propose. I shall be careful to indicate what are internal and what are external critiques.

I now reiterate, in greater depth than previously in subsection 3.4, and sections 2 and 8, how Rawls's position on the scope of justice altered throughout his career. In the beginning, possession of the capacity for a sense of justice is necessary and sufficient for one to be owed justice.[475] However, come *A Theory of Justice*, the capacity for the powers is no longer taken to be necessary, but is merely sufficient. What are the other sufficient criteria? Rawls does not say directly, but the options can be reconstructed, particularly in the light of what comes next in *Political Liberalism*. There, the criteria for one to be owed justice are the capacity for the two moral powers *and* the capacity to cooperate directly in the maintenance of the basic structure of society. In addition, by *Political Liberalism* and the later works, it is no longer clear that the capacity for the two moral powers is sufficient in itself. It may be, however, given what is said in *A Theory of Justice*, that the capacity to cooperate *is* sufficient. Hence Rawls's final position most likely appears to be that the capacity to cooperate is sufficient to be owed justice, but the capacity for the two moral powers is not sufficient by itself. I view this position as untenable — even by Rawls's own lights. Hence, I aim to defend the original formulation as given in "The Sense of Justice" as the best account of the scope of justice available for the Rawlsian contractarian.

A note on the status of the chapter within the context of the rest of the thesis. Previous chapters have largely concentrated on exegesis, making only small gestures towards substantive criticism. This chapter is different. It first presents arguments for the revision of Rawls's account of the scope of justice in line within his earliest position. These arguments are heavily critical of certain later elements of Rawls's theory, and possibly go beyond what he would himself have been happy to accept. The arguments are hence not simply arguments from within Rawls's overall theory, in its various forms, but also contain considerations of my own. This discussion, which takes up most of the chapter,

475 As is, we can assume, the ability to develop a conception of the good. This second power is not explicitly mentioned in the original article.

elaborates a certain contractarian account of the scope of justice. However, in section 18, my criticisms are then extended to contractarian accounts of the scope of justice themselves.

My argument for turning the clock back to the original specification of the scope of justice, and then criticising even that specification, will proceed as follows. I take issue with the position put forward in *Political Liberalism* and the later philosophy more generally. First, in sections 15 and 16, I argue that the ability to contribute to a society is irrelevant to being owed justice. In section 15, I argue that possession of the capacity for the two moral powers is always sufficient, and that the ability to contribute is not necessary. In section 16, I argue that the ability to contribute to the basic structure of society is never sufficient by itself. In section 17, I address whether there are any other sufficient criteria apart from the moral powers. I claim there are not. Sections 15, 16 and 17 together entail that the capacity for moral powers is necessary and sufficient. My final position, then, is that Rawls should say that when beings owe each other justice, they simply must be capable of developing the moral powers. In section 18, however, there is a concluding reflection relating to whether we should hence endorse Rawls's account of the scope of justice. I believe that we might, but that there appear to be to me problems with Rawls's position overall, which stem from combining this account of justice's scope with the ideas of political liberalism, public justification and legitimacy.

Section 15: The Moral Powers and the Ability to Contribute

15.1 Society as fair cooperation, and justice as reciprocity

This section proceeds as follows. This subsection, after pinpointing the specific focus of the section over all in contrast to sections 16 and 17, sets up Rawls's account of the moral powers as the basis of equality, and then introduces the ideas of society as fair cooperation, and justice as reciprocity. Subsection 15.2 then investigates just what Rawls conceives to be the product of fair social cooperation in the well-ordered society of justice as fairness. Subsection 15.3 introduces the problems which have been raised with the idea that the moral powers and the ability to contribute to the productive scheme of society are necessary and sufficient to be owed justice. Subsection 15.4 argues for the idea that the capacity for the moral powers should be sufficient in itself to be owed justice, and that the ability to contribute should not be necessary. Subsection 15.5 then lays out an extensive

host of objections and queries regarding whether this revision is legitimate, and how much it may alter the structure of Rawls's theory.

To begin: at various points, Rawls appears to commit himself to the idea that the capacity for the moral powers, and the ability to cooperate in society in certain ways, are jointly necessary and sufficient to be owed justice. The most overt statements are found in *Political Liberalism* and *Justice as Fairness: A Restatement*.[476] I say "appears", but for the most part I shall assume that Rawls does make this commitment. Only in subsection 15.5 A do I address whether this is strictly correct. In any case, I thoroughly reject, over the course of this section and the next, what I shall often call "the contribution requirement".[477] This is the requirement that it is necessary, or alternatively sufficient, that an individual contribute to the cooperative surplus of a cooperative scheme in order to be owed justice. Contribution is sometimes given to be a necessary condition – along with possessing a capacity for the moral powers – to be owed justice. This is the formulation which this section (section 15) will consider. Alternatively, other passages taken together suggest that contribution is a sufficient condition, by itself, to be owed justice. This idea will be tackled in section 16. The eventual conclusion of this section (section 15) is that the capacity for the moral powers is sufficient by itself to be owed justice. This section (section 15) hence serves to reject the claim that a capacity for the moral powers, and contribution to the cooperative scheme of society, are jointly necessary and sufficient. Contribution is not necessary, and the capacity for the moral powers is sufficient in itself. It is only in the next section (section 16), however, that I argue that contribution in itself is not sufficient. And it is only in section 17 that I argue that there are no other sufficient criteria to be owed justice other than possessing the capacity for the moral powers. Taken together then, sections 15, 16 and 17 entail that the capacity for the moral powers is both necessary and sufficient to be owed justice.

I now introduce Rawls's account of the basis of equality, and the variability he allows in people's capacity for the moral powers. I will subsequently introduce the idea of society as a fair scheme of cooperation over time, and the idea of justice as reciprocity.

The two moral powers, previously discussed in sections 2 and 8, and subsection 3.4, are the sense of justice, and the capacity to develop a conception of the good. It is obvious that people vary in the degree to which they develop these capacities. Some people may

476 See, for example, *PL*, pp. 15—22, *JF*, p5—8. The idea is at play in *Theory* as well (for example pp. 4/4, 84/73—74, 88—89/76—77) though it is not explicitly mentioned in the discussion on the basis of equal justice in §77. For ease of exposition, I have assumed that *Theory* alters the account in "The Sense of Justice" only in making the moral powers sufficient instead of necessary and sufficient, and does not add the requirement to be able to cooperate. If this is strictly incorrect, my arguments are in any case unaffected.

477 This terminology follows Vanderschraaf (2011)

excel in organising the achievement of their life's ambitions, whilst others may be better or more easily able to comport themselves in a just manner.[478] As noted in subsections 3.4 and 8.3, the capacity for the moral powers forms the basis of moral equality in Rawls's theory. The criticism might arise that, with such individual variety, the moral powers cannot serve as the basis of the most basic kind of equality.[479]

Rawls rejects this criticism. The most basic kind of equality is equality of fundamental respect or recognition, which "is owed to persons irrespective of their social position."[480] Basic equality is based simply on the *possession* of the capacity for the moral powers to some minimum degree. Interpersonal variation in the realisation or capacity to realise the two moral powers is irrelevant to basic equality. Once the minimum conditions for the moral powers are met (either contemporaneously or prospectively),[481] then an individual is not only owed justice, but equal justice. The moral powers are hence said to constitute a "range property" which marks out those who deserve just treatment. People can fall within a certain range of varying moral psychological characteristics. Nevertheless, they can still be said to equally possess the property of having the moral powers – of falling within that range.[482]

Picking out a range property is essential. The simpler precept of treating equal cases equally will not do alone. With that precept "there is no guarantee of substantive equal treatment, since slave and caste systems may satisfy this conception."[483] Justice could require more and/or deliver more to those whose moral powers were more developed. Justice for Rawls does make different demands on different people, but this is a function of their social position in a just institutional order, not their basic status in the eyes of justice.[484]

Precisely what constitutes the minimum is held by Rawls to be, to some extent, irreducibly vague. He writes

> The conception of moral personality and the required minimum may often be troublesome. While many concepts are vague to some degree, that of moral personality is likely to be especially so. But these matters are, I think, best discussed in the context of definite moral problems. The nature of the specific

478 See *TJ*, p. 506—507/443
479 *TJ*, p. 507/444
480 *TJ*, p. 511/447
481 *TJ*, p. 509/445—446
482 *TJ*, p. 508/444—445
483 *TJ*, p. 507/444
484 *TJ*, p. 511—512/447—448. For a recent defence of this approach to equality, see Carter (2011)

> issue and the structure of the available general facts may suggest a fruitful way to settle them.[485]

The rest of section 15 assumes that persons are above the minimum threshold, whatever that might be, and considers whether a further condition should be relevant to someone being within the bounds of justice. This further condition is that, as well as possessing the moral powers, a person should be able to contribute to the maintenance of the basic structure of society.

This further condition stems from two related fundamental ideas found in Rawls's work. For Rawls, the basic concern of a theory of justice is society conceived "as a fair system of cooperation over time, from one generation to the next."[486] A scheme of social cooperation is one in which everyone is benefited by everyone playing their part in shared rules of cooperation.[487] Cooperation creates an infrastructure for the distribution of a supply of goods, to which each contributes and from which each benefits (see earlier section 8.2).

This idea is the most fundamental in Justice as Fairness. The ideas of the moral person or citizen[488] (characterised by the possession of the capacity for the two moral powers), and the well-ordered society, are both characterised by reference to it.[489] Hence, sacrificing this understanding of society would not be a small matter for Rawls (see subsection 15.5 L below). Speaking of society as a fair scheme of cooperation over time is a bit of a mouthful. Hence, I shall from now on use the locution "society as fair cooperation".

The idea that justice concerns society as fair cooperation is closely linked by Rawls to the idea of justice as reciprocity. This conception of justice is said to occupy the centre ground between two others: justice as mutual advantage, and justice as impartiality. The terms were coined by Brian Barry to describe the following two positions. In justice as mutual advantage, "the just terms of cooperation are those that would have been agreed upon by people [merely] trying to do the best for themselves" if they were situated at "a non-agreement point from which the hypothetical bargaining is to start."[490] In justice as impartiality, by contrast, just arrangements must correspond to "what can be approved of

---

485 *TJ*, p. 509/445
486 *PL*, p. 15. See also *TJ*, pp. 4—5/4—5 and *JF*, p. 5
487 *PL*, p. 16, *JF*, p. 6
488 "Moral person" is the relevant concept in the earlier, comprehensive statement of Justice as Fairness (*TJ*, p. 505/442), "citizen" in the later, political version (*PL*, pp. 18—20, 29)
489 *PL*, p. 14. See also *JF*, pp. 24—26
490 Barry (1989) pp. 367—368. See also pp. 5—7, 359—361

from an impartial standpoint."[491] Justice as mutual advantage conspicuously excludes those unable to contribute to cooperative schemes from moral concern all together. People's places in the agreed cooperative schemes are determined on the basis of their bargaining power, so if you have no bargaining power you will also be excluded from any protection.[492] Given that your power determines your basic position, justice as mutual advantage does not conform to our intuitive concepts of fairness and reasonableness, but only to rationality.[493] Justice as impartiality, by contrast, simply includes all those with interests — which includes all humans, at the very least.[494]

Barry held that Rawls's own theory awkwardly incorporated both elements of justice as impartiality and justice as mutual advantage.[495] Adopting a suggestion made by Allan Gibbard in a review of Barry's work,[496] Rawls replied that Justice as Fairness was actually a member of a third intermediary view - justice as reciprocity.[497] Justice as reciprocity holds that justice requires fair mutual advantage between persons, arising from mutual reciprocation within fair institutions. This mutual advantage is fair in that what persons receive is not conditioned by the quantity or quality of goods they are able to contribute to society, and certainly not by the threats they are able to bring to bear on others.

Rather, what you get is what could be reasonably agreed by all as increasing the prospects of everyone who is cooperating, starting from a baseline of equal shares in the cooperative surplus.[498] To get this baseline, we assume that the proceeds of social cooperation – in Rawls's case the social primary goods (see below) – are going to be shared out equally between all social positions.[499] Shares may then legitimately become unequal only if inequalities would serve to raise the absolute shares of everyone, including the worse off.[500] You yourself may be able to contribute very little to this surplus. But even if this leads to you occupying the least favoured social position, you will receive a great

---

491 Ibid. p. 362. See also pp. 7—9, 284, 361—363

492 See, for example, ibid. p. 249

493 See *PL*, p. 48. See also Scanlon (1998) pp. 191—197

494 I take "having interests" to be the easiest way to summarise Barry's many distinct discussions of who is included within the scope of justice, and morality more generally. Barry holds that at least some of the provisions of justice extend to all humans, even if severely disabled, (see Barry (1989), pp. 244—254, and (1995a), pp. 42—43, 60). The rider "at least" follows from the fact that non-human animals are also seemingly included (See Barry (1995a), pp. 86, 208)

495 See Barry (1989) chapters 5 and 6

496 Gibbard (1991) esp. pp. 266—273

497 *PL*, p. 17 fn 17. Note that Rawls does not expressly use the phrase "justice as reciprocity," as Gibbard and Barry do. Nevertheless, he conceives of the ideal of justice as a fair reciprocal relationship between agents. Hence, the phrase isn't misleading.

498 *PL*, pp. 16—17, *JF*, p. 6. For the more precise reasoning for starting from an equal division of the cooperative surplus, see *TJ*, pp. 101—104/87—90, *JF*, pp. 74—77

499 See *TJ*, p. 62/54—55, 150—151/130, and also *LP*, p. 41

500 For example, see *TJ*, pp. 60—65/52—56, 151—152/130—131, and *JF*, pp. 61—64

deal more than what you put in. Those who can contribute greatly, and who have significant bargaining power and threat-advantage, may be net losers in benefits, in contrast to a society in which their shares are linked to their bargaining advantage. But justice requires rejecting societal schemes which allow bargaining advantages to play a part in determining the benefits of social roles. One's ability to contribute should be seen as a morally irrelevant factor for deciding cooperative shares; one's ability to threaten is a morally unacceptable one.[501]

However, in justice as reciprocity, contribution to the cooperative scheme of society is still required to be included within the remit of justice. Justice requires a productive relationship of reciprocity – of both sides fairly benefiting from each other[502] – and hence justice as reciprocity falls short of the scope of justice as impartiality. Justice is based on reasonableness for Rawls (section 8.2), not on altruism, or some mixture of the two.[503] Barry disputes whether such a middling position can at all be coherently maintained;[504] I will return to this in subsection 16.5 P.

The ideas of society as fair cooperation, and justice as reciprocity, exclude a certain class of beings – those we shall call the *non-contributors* – from justice. Just who the non-contributors are can be extensively investigated and debated. Candidates often proposed include the congenitally impaired and chronically ill, the people of the future, and animals.[505] That members of such groups should be excluded from justice on the basis of their inability to cooperate in society as a fair scheme of cooperation has been frequently criticised.[506] In my discussion, I shall mainly be considering the physically and mentally impaired[507] — all the time specifying, throughout this section, that they possess the capacity to develop a sense of justice and the ability to develop a conception of the good.

---

501 See *TJ*, pp. 102—105/88—90, 133—134/115—116, and also *JF*, pp. 72—77

502 *JF*, p. 61 expresses this particularly strongly.

503 As we might reconstruct positions offered by Barry (1995a), Stark (2009)

504 See Barry (1995a), pp. 46—61

505 Related issues arise relating to the periods in everyone's lives when they are unable to contribute: when children, ill or elderly. But Rawls has answers regarding these groups – though I shall not here address how plausible they are. On children see *TJ*, pp. 462—467/405—409, 509/445—446. On health care see *PL*, pp. 183—186 and *JF*, pp. 171—176.

506 For example, Barry (1989), pp. 234—249, (1995), pp. 59—60, Nussbaum (2006), pp. 22—25, 56—67, 107—154, 330—338, Kittay (1999), chapter 4

507 Martha Nussbaum relates that impairment and disability are defined in the disability literature in the following way. An impairment is "a loss of normal bodily function", a disability "is something you cannot do in your environment as a result." See Nussbaum (2006) p98 fn 5. Impairments need not always lead to disabilities. The arrangement of one's environment may or may not lead to a disability, depending on whether your impairment is taken account of by those who arrange your environment. I note that a wider and more formal notion of impairment can be specified, which simply states that impairments are lacks of possible bodily functions. Hence, an impairment of mine is that I do not possess a system of echo-location: an impairment not shared by a bat. This need not be taken to imply that I am "impaired", in a looser, colloquial sense. This usage seems to fit better with Nussbaum's repeated insistence that those we routinely identify as "disabled" persons should not be seen as aberrations of "normal" persons, i.e. pp. 99, 101. No doubt many complications would arise from this revised usage – complications I do not consider here.

These persons represent the clearest test case. In section 16.5 M below, I shall comment briefly on how the position I come to relates to animals and future people. It is important to note that Rawls only excludes these groups from justice due to their lack of ability to cooperate, and not simply due to the fact that they are impaired *per se*. It is also important to recognise that the classes of those recognised as impaired and/or disabled, and the non-contributing, are distinct. I am in the process of defining the latter philosophically, whereas the former will here be left as a broad and fuzzy folk-category.[508] Many of those we recognise as having an impairment – the blind, for example – have for many years contributed to societies in the ways Rawls requires. However, we have yet to specify just what this relevant type of cooperation is. The next subsection will do so.

15.2 What sort of cooperation? To produce what?

In this subsection, I first characterise the idea of contributing to a cooperative scheme in an abstract manner. To then describe what Rawls understands contributing to a cooperative scheme to be, I then first describe Rawls concept of the social primary goods, and then second describe what Rawls counts as contributing to the cooperative surplus of those goods, and hence what counts as meeting the contribution requirement. Subsection 15.3 then briefly observes how others have attempted to defend Rawls's contribution requirement, and also emphasises the distinction between possessing the two moral powers, being able to contribute to a productive cooperative scheme, and being physically or mentally disabled and/or impaired.

To begin speaking abstractly, to contribute to a cooperative scheme is to produce a good through following the rules of that scheme. The good may then go on to benefit others, or oneself and others. But, importantly, the good goes through the institutions first, given that we are talking about schemes of cooperation, rather than just cooperation *per se*. Cooperative schemes cannot exist between individuals none of whom are able to benefit each other in any way. Whatever else must be assumed for there to be cooperation, there must be at least two individuals benefiting each other. Cooperation must also consist in more than this. I benefit from the natural world around me, and at some point it will benefit from me, even if only when I'm dead in my grave. This does not intuitively amount to me cooperating with the natural world. Cooperation requires jointly recognised intentions connecting the cooperating agents – if not everyone to everyone else, then through chains

508 This is not to say that it *should* be so left. Philosophical work on disability gives us much reason to be suspicious of our folk conceptions on this topic. But I cannot engage directly with this literature here.

of connection covering all those involved. In summary, "cooperation is guided by publicly recognised rules and procedures that those cooperating accept and regard as properly regulating their conduct."[509]

What if a group of individuals benefit each other in a shared practice, but in addition benefit a non-contributing third party? It seems right to say that this outlier isn't cooperating. Certainly they do not contribute. But are they included within the cooperative scheme? This seems less cut and dried. They benefit from it. There is nothing to suggest that there can't be cases where the non-contributor understands it, and can convey that they understand it. To take an extreme case, imagine a paralytic who is nevertheless able to blink his eyes to indicate "yes" or "no" to questions. It seems acceptable to say that they can be part of the practice, and even have a position within it. What needs to be stressed is that they can play no role in sustaining it in the productive form that it has. More needs to be said, but this seems sufficient to head off a suggestion that an inability to contribute to a set of cooperative institutions means that, logically, one cannot be a part of such institutions. Cooperative institutions need not hold exclusively between cooperative agents.[510]

We have so far left open what exactly the contributors in society as fair cooperation contribute to. The basic subject of justice for Rawls – what persons in a modern society cooperate to produce, and reproduce – as introduced in section 2 is that society's basic structure of institutions.[511] The basic structure serves to distribute the primary social goods.[512] These goods were previously mentioned in 9.1. The primary goods are those goods any person is rationally presumed to want whatever else they want.[513] Primary goods are social when they are directly under control of the basic structure, whereas primary goods considered more generically can potentially be only indirectly influenced by that structure.[514]

In later work, the social primary goods are linked more closely to the conception of the person in Justice as Fairness (sections 8 and 9). Rather than simply being the social primary goods any person is rationally assumed to want, they are the social primary goods

509 See *PL*, p. 16

510 Silvers and Francis (2005) make a similar point, and develop its implications into an argument for including the non-contributing within a broadly contractarian theory of justice. See, in particular, pp. 45, 68—73. I believe that their position is compatible with Rawls, but cannot be substituted into the foundations of his theory without leading to wider revisions of his basic ideas than I here propose, particularly to the original position.

511 See *TJ*, pp. 7—11/6—10.

512 *TJ*, p. 62/54—55. Rawls refers interchangeably to "social primary goods" and "primary social goods", though these might be thought to have different connotations. I follow him here.

513 *TJ*, p. 62/54—55, 92/79

514 *TJ*, p. 62/54—55

"normally needed for developing and exercising the two moral powers and for effectively pursuing conceptions of the good with widely different contents."[515] I have assumed this revision throughout the thesis.

The social primary goods are listed as rights and duties, income and wealth, and the social bases of self-respect.[516] As they are to be distributed through the basic structure, which is governed by a public conception of justice, the social primary goods and their distributions are assumed to be capable of being publicly observed and accounted for.[517]

The social positions in society that are recognised by the institutions of the basic structure are presumed to be those of a person cooperating in the overall maintenance of that structure throughout a full life. We can gloss this and say: they are the social positions of the people who, for most of their lives at least, have some kind of job within either the private or the public sector.[518] The least well-off position in society is that of the employed person with the lowest expectations of the social primary goods.[519] Note also that some of the social primary goods, such as the basic liberties and the means to use them, and the social bases of self-respect, must be distributed equally.[520] The least favoured position does include, along with all other social positions, provision for illness and temporary disability. But this is justified by reference to the need to enable each person to return to their place in society, and their work, in the event of illness or some other misfortune.[521] The provision is not conceived to have itself some special or lexical weight, or primacy — healthcare is to be balanced against other competing distributive demands as required by the two principles of justice overall.[522] Nor is healthcare conceived to be distributed on any other basis other than the need to maintain a person's use of their moral powers, at least in the eyes of justice.[523]

Given all these features of the account of the basic structure and social primary goods, what contributing to society amounts to in Justice as Fairness is seemingly taking up employment in the system of institutions governed by the basic structure. Through this, persons contribute to the upkeep of the basic structure and the distribution of primary social goods which flow through it. This implies that if one is completely, and not just

---

515 *PL*, p. 76. See also *CP*, pp. 312—313, 365—366, *JF*, pp. 57—59.
516 See *TJ*, pp. 62/54, 92—94/79—80, *PL*, p. 181, *JF*, pp. 58—59.
517 See *TJ*, p. 95/81, *PL*, pp. 181—182, *JF*, pp. 59—60, *CP*, pp. 363—364
518 This is supported by Rawls's comments that the least-well off are not to be defined as those reliant on state welfare. See *JF*, p. 138—140, 179
519 See *TJ*, pp. 93—94/80. See also *JF*, pp. 59—64
520 For example, see *TJ*, p. 93/80. I say that the social bases of self-respect must be distributed equally, assuming that arguments given by Eyal (2005) p. 197 that this is the right reading of *TJ*, p. 546/478—479
521 *JF*, pp. 173—175, *PL*, p. 184
522 See *JF*, pp. 173—174
523 On this last point, see comments at the start of the next subsection, and also below in subsection 16.5 A and section 19.

temporarily, unable to take up employment, one does not adequately give to society's fair scheme of cooperation. One is then outside the scope of justice.

15.3 Moral Powers, and ability to contribute

As I have mentioned, this restriction on justice has been widely criticised. Several defences of Rawls's theory on this point have also been proposed. They aim to show either that according to Rawls, those unable to contribute to society in this way are not excluded from justice entirely,[524] or else are nevertheless shown adequate moral concern under other duties and obligations in Rawls's scheme.[525] Others have proposed alterations to Rawls's theory in order to deal with the issue.[526] What does not seem to have been adequately recognised by most of this literature, however, is that possession of the two moral powers, and the ability to cooperate, are not coextensive. A person can possess a sense of justice and an ability to develop their own conception of the good, without being able to take up a position within the basic structure of society. Conversely, a person lacking a sense of justice can still take part in cooperative schemes. Many writers either fail to observe the first fact at all, or else fail to observe it in a sufficiently systematic way.[527] What is required is for the ability to contribute, capacity for the moral powers, *and* being impaired or disabled all to be clearly distinguished.[528]

That the moral powers can obtain without cooperative ability must be the right interpretation.[529] Rawls's phrasing indicates that he recognises a distinction between the possession of the moral powers and the ability to employ them in cooperative ventures.[530] Falling below the minimum needed to cooperate can be in terms of either "moral, *intellectual* or *physical* capacities."[531] A person who is in traction for several months is unable to go to work. But we would not say that that person, for that period, lacks a sense of justice, or a capacity to develop a conception of the good. Hence, we would not say the

524 For example, Freeman (2006) pp. 411—418

525 See Kelly (2010) pp. 63—66. Quong (2007), pp. 91—97 aims to include the non-contributing within justice under the aegis of the natural duty to mutual aid. I do not believe this duty is a duty of justice for Rawls. He appears to clearly distinguish the duty from that of justice (*TJ*, pp. 333—339/293—298, 511/447). But this would still allow the non-contributing some level of moral consideration – in certain respects a quite significant level if Quong's argument is sound. I argue in subsection 15.5 A that being the subject of these natural duties is still inadequate recognition, however.

526 For example, Richardson (2006), and Stark (2007)

527 For example, Nussbaum (2006) chapter 2, Freeman (2006), Richardson (2006), Quong (2007)

528 Stark (2007) pp. 129—132 explicitly discusses these distinctions. Discussions by Stark (2009) pp. 80—81, Kelly (2010) pp. 63—66 and Terzi (2010) pp. 155—161 also make explicit use of them. My eventual position differs from each of these writers.

529 Contra Nussbaum (2006) pp. 127—135

530 See, for example, *PL*, p. 19

531 *PL*, p. 184 my emphasis. Stark (2007) p. 130 makes precisely the same point.

same for someone who is permanently physically disabled to the same degree. And we can say the same for certain mental disorders or cognitive impairments, whether temporary or permanent.

Furthermore, people who lack a sense of justice – the purely self-interested – can still be willing and able to cooperate in fair cooperative schemes, at least when it is to their advantage to do so. As Brian Barry writes "so long as even very rough equality of strength obtains among the parties to rules of justice, the rules recommended by justice as mutual advantage will tend to correspond to those that we would ordinarily think just."[532] It might be wondered whether the actions of such people count as genuine cooperation, given our specification above in subsection 15.2. Similarly to what was said earlier about whether those unable to cooperate can be part of cooperative schemes, we might say that beings need not fully share the intentions behind a scheme to be said to take part in it. For the self-centred surely share some of the potential and acceptable motives of the rest, given that the scheme is mutually advantageous. I am unsure whether this reply is fully adequate. But for the purposes of this discussion I shall assume that it is, or else that those lacking a sense of justice can in some way be correctly said to cooperate. Whether the ability to cooperate in those who cannot develop the moral powers is sufficient to be included within the scope of justice will be taken up again in section 16.

15.4 Contribution is not required for justice

Given that the moral powers and the ability to cooperate come apart, are both of these necessary (and also jointly sufficient) to be owed justice? I maintain that those possessing the moral powers, or their capacity, but lacking the ability to cooperate in the maintenance of the basic structure, should be unambiguously included within the scope of justice. This means that contributive potential is not necessary to be included within the scope of justice, and the contribution requirement, construed as a necessary requirement, is misplaced. I also maintain that Rawls should have said that the possession of the capacity for the moral powers is sufficient, and that he does not do so leads to internal tensions within Justice as Fairness. I also present independent moral reasons for holding that moral power capacity is sufficient.

I first defend the basic idea that possession of the moral powers, or the capacity to develop them, is sufficient for justice. These are independent moral reasons – external to

532 Barry (1995a) p. 45

Justice as Fairness. Next, I shall highlight those elements of Rawls's work which support the position I am defending. This subsection, then, concentrates on laying out my basic case. The exposition will raise a number of questions for any astute reader. I attempt to answer a host of them in subsection 15.5.

We can most vividly see the problem with not including non-contributing beings who have or can have the moral powers within the bounds of justice by considering the implied attitudes and perspectives of the members of the well-ordered society.[533] What we postulate as moral ideals, these persons will fully psychologically realise in their thoughts, feelings and deeds (subsections 3.2, 9.2). Hence, they will embody the contribution restriction in their thoughts and feelings. If we accept the stipulation that the ability to cooperate in the maintenance of the basic structure, by accepting some kind of employment, is necessary to be owed justice, then in a well-ordered society people will not see justice as owed to those who cannot so cooperate. This attitude will be shared by all persons with the moral powers: both those who can *and* those who cannot cooperate. A striking feature of this society is that this will be the case *even though* individuals in the latter group may potentially have a better understanding of the rights and duties of the just person or citizen, and/or have greater motivation to defend and serve as an advocate for those rights and duties, than those who are able to contribute to the upkeep of the basic institutions.

If they genuinely do possess the two moral powers, those unable to contribute to the basic structure through that structure's recognised positions will nevertheless be motivated to uphold the justice of society. And surely this, if anything, expresses good-will towards their fellows. The inability to take up some role in the economically productive arrangements of society is merely due to some kind of disability or impairment, either physical or mental. But it is on the basis of such impairments that the *non-contributing themselves*, as well as everyone else, will accept that the non-contributing cannot be granted basic justice.

The non-contributing members of society can hence be thought to see things like this: “Because we are unable to help in maintaining the basic institutions of our society, we cannot be granted what is owed to someone who does play a part in maintaining them. It doesn't matter that we are willing to, and that it is only through some misfortune that we are unable to. It would be grossly unfair to grant us any of the provisions *or recognition* which come from being included within the scope of justice, when we are unable to

533 The kind of approach I adopt here is inspired by the kind of “interpersonal test” proposed by Cohen to test mooted principles of social organisation. See Cohen (2008) pp. 35—48. It works slightly differently, by assuming an ideal society which thoroughly adopts the proposed principles, and then considering what our moral intuitions are regarding such a society.

contribute to the cooperative surplus."

Spelt out like this, the requirement begins to seem deeply problematic. The last sentence brings out what is wrong. People who have the moral powers necessarily *are* able to give basic recognition to others who have or are capable of having the moral powers. And as we saw above in subsection 15.1, Rawls holds this to be the most basic sense of equality from the perspective of justice. This equality is meant to serve as the basis for the *rest* of the rights and duties of justice. If we assume that all the requirements of justice must be founded on the basic equality of recognition, then it does not seem that we can ever exclude a being with a sense of justice from the scope of justice, whatever their cooperative potential. Basic recognition brings with it a stake in the cooperative surplus, however this is ultimately to be divided up.[534]

In summary, intuitively it does not seem that we can approve of the attitudes of a society of persons who did place this restriction on the scope of justice. If a person has a sense of justice, then they are able to give basic recognition to those who are similarly endowed. They should hence be owed the protections (and have the responsibilities) of justice. Other facts about their person, such as possessing certain disabilities or impairments, should not be relevant. But the contribution requirement makes them relevant. Hence the contribution requirement is morally suspect.

These are external moral reasons for rejecting the contribution requirement, and accepting that the capacity for the moral powers is sufficient to be owed justice. Various aspects of Rawls's work indicate that this is what he should have maintained. Some of these have been remarked upon by various philosophers. Brian Barry notes that restricting the scope of justice on the basis of physical or intellectual, but not moral, privations offends against one of Rawls's basic moral assumptions.

> Natural and social advantages that make people more of less productive are a matter or good fortune and hence do not constitute ground-floor claims to receive more or less of the social product. This notion, however, clearly implies that the congenitally disabled cannot be held responsible for lack of productivity and should therefore have a valid claim on a share of their

534 Several writers have made the point that the disabled are able to give basic recognition to other beings with a sense of justice. See, for example, Nussbaum (2006) pp. 121—122, 128—130, 133—135 and Silvers and Francis (2005), p. 68—73. As mentioned above, these authors do not precisely specify that we are talking about persons possessing, or able to possess, the two moral powers, who are unable to take employment in the basic structure of society. In addition, we are not talking about any old kind of recognition, but specifically the recognition that a person has or is capable of having the two moral powers, which is not necessarily the same as the recognition talked about by Nussbaum and Silvers and Francis.

society's resources.[535]

No philosopher to have engaged in such internal critiques,[536] however, has explored the aspects of Rawls's moral psychology which tell against the contribution requirement. This is despite of the fact that Rawls's initial discussions of the basis of equality – in "The Sense of Justice" and *A Theory of Justice* – are both found situated within broader discussions of that psychology.

There are several aspects of the moral psychology which seem incompatible with the contribution requirement. Rawls holds that one of the factors involved in our development of the sense of justice is the recognition of "an unconditional caring for our good."[537] This unconditional care is presumed to start from birth, if not before. It is directly evidenced by our parents, and indirectly, through them, from the society around them.[538] This care is not unconditional in every sense. Moral conduct is eventually expected of a person as they develop, as opposed to what we might call perfectly altruistic care.[539] But it is unconditional in that the person receives care for their own sake, and not simply as a means to something else.[540] Now persons with a capacity to develop the moral powers can ideally be expected to receive this care. What loving parent or guardian wouldn't express both altruistic and non-altruistic (what we might call reciprocal) care towards their children?[541] As the child grows, however, and is seen as capable of the two moral powers, then altruistic care becomes less and less appropriate. More and more is expected of them, and reciprocal care is repeated stressed. Such attitudes would not be withheld simply on the likelihood, or even the certainty, that their child would never be able to join the workforce of civil society.

Again, ideally, the attitude of such child-rearers would be mirrored in similar attitudes and institutions of society at large. Though the sentiments of love (and friendship), and justice for Rawls are distinct, they are continuous with one another.[542] The love between a family, or friendship between similarly close associates, lays the foundation for

---

535 Barry (1995a) p. 60

536 In certain ways, Richardson (2006) and Stark (2007) can also be taken to be at least partially internal critiques

537 *TJ*, p. 498/436. See futher Appendix II.

538 *TP*, pp. 464/406—407, 473—474/414—415, 490—491/429—430

539 *TP*, pp. 466/408, 498/436

540 See *TP*, pp464—465/406—407, 499/436

541 This statement appears to be unproblematic given the moral psychology as it is presented in the early philosophy. However, certain features of the later philosophy, when combined with the quite minimal content of the conception of the person outlined in sections 7 to 9, may cause problems for this intuitively commonsensical assertion. I do not explore this issue here, and point out the disparity between Rawls's original psychology and the contribution requirement.

542 *TP*, pp. 476/417, 478/419

attitudes of justice towards wider society. Rawls's principles of moral psychology posit that the attitudes expressed towards us by successive and expanding circles of relationships – family, private and public associations, society in general – bring about the development of the sense of justice.[543] If these attitudes cease to extend once we are faced with someone who cannot cooperate in the maintenance of the basic structure, there needs to be an explanation for this. *Prima facie,* the continuity of the sentiments of love and justice implies that such persons will also be included within the scope of justice. To posit otherwise would require a disjoint in Rawls's principles of moral psychology. It would have to be that the parents or guardians of those with the capacity for the moral powers, but without cooperative ability, show the appropriate attitudes of reciprocating care towards them, as to their friends. But then wider society does not express those same attitudes, because wider society, in addition, requires that persons be able to contribute to the maintenance of the basic structure of society.

We would have to say that further conditions for being afforded the responsibilities and protections of justice are recognised, consisting in meeting certain physical and mental requirements which are distinct from simply possessing the sense of justice. What opposing sentiment would these correspond to? The person unable to meet such requirements is no longer recognised for their own sake solely on the basis of their ability to recognise (and comply to the extent that they can with) fair moral requirements. The unconditionality of their parent's or guardian's care for them is not reflected in society's attitudes towards them. In the well-ordered society, where the ideal of justice as reciprocity is embodied perfectly in its member's psychology, everyone will accept that society need not have an attitude towards such persons that is continuous with the sentiments those persons' parents have towards those persons. People in general need not value them intrinsically, on the basis of their just sensibility at least. But, by implication, the able-bodied people in such a society do not value each other solely on this basis either, but conditionally on their ability to contribute. The progression of unconditional sentimental attachment posited by Rawls in the well-ordered society is hence broken by the contribution requirement. This is sufficient to render that requirement incompatible with his moral psychology.

Problems can be found at an even deeper level. The principles of moral psychology are said to be *based upon* the idea of reciprocity: "a tendency to answer in kind"[544] (see further Appendix II). Children answer in kind to their parent's or guardian's love when they

543 See *TP*, pp470/411—412, 473—474/414—415 490—491/429—430. See also Appendix II
544 *TP*, p. 494/433

love them in return and live up to the justified standards they impart. Such attitudes are continuous and eventually develop into a sense of justice. When grown, such persons will recognise others who possess or are capable of possessing a sense of justice, and will both respect and have expectations of them. This moral development embodies the idea of reciprocity. But the question then becomes – why does a seemingly more demanding standard of reciprocity appear by the time we get to the basic structure of society? The idea of reciprocity, as a psychological tendency, does not appear to make such a distinction itself (see Appendix II). If the moral sentiments and natural attitudes are, ideally, continuous with each other, and this is an appropriate specification of the moral principles and ideals of Justice as Fairness (see subsections 5.2 and 9.2), there seems no reason to construe the well-ordered society of Justice as Fairness's conception of reciprocity in a compartmentalised way: as requiring *at bottom* something different at the level of society's basic institutions than in the associations of civil society. I say “at bottom” here in order to leave open the question as justice does require different things from different social roles. But this variation is not meant to alter basic equality.

These central aspects of Rawls's moral psychology fit ill with society as fair cooperation, and justice as reciprocity — at least, as they are specified. This is in addition to the independent moral arguments which I presented earlier. The fundamental point is that it is only the possession of or capacity for the moral powers which is meant to be relevant from the point of view of justice. As we have seen, possession or capacity can come apart from cooperative ability. Absent some further argument, cooperative ability remains morally arbitrary.

However, it might still be thought that the proposal to drop the contribution requirement is nevertheless unnecessary, or else would require such wide reaching revisions of Rawls's theory as to substantially change its character. In the next subsection I address a variety of such claims.

15.5 Possible objections

I reject society as fair cooperation and justice as reciprocity as Rawls formulated them. It may be argued I have done so erroneously. I will argue that I have not. Further it may be thought that this implies the rejection or alternation of many other aspects of Rawls's theory. I will argue either that it does not, or in the cases that it does, these alterations are acceptable.

I lay out each potential question or criticism in turn in the following flurry of

subsections. I shall not summarise them here. Instead, each is prefaced by an italicised question or statement which the following subsection should be taken to address. Where these subsections are linked to each other, this will be indicated.

15.5 A: Rawls understands Society as Fair Reciprocity as an ideal of justice, not as a limit on the scope of justice. Hence, he does not place the non-contributing outside the scope of justice.

This idea is suggested by passages such as this one

> Since we begin from the idea of society as a fair system of cooperation, we *assume* that persons as citizens have all the capacities that enable them to be cooperating members of society. This is done to *achieve a clear and uncluttered view* of what, for us, is the *fundamental* question of political justice: namely, what is the most appropriate conception of justice for specifying the terms of social cooperation between citizens regarded as free and equal, and as normal and fully cooperating members of society over a complete life?
>
> By taking this as the fundamental question we do not mean to say, of course, that no one ever suffers from illness and accident ... But given our aim, *I put aside for the time being* these temporary disabilities and also permanent disabilities or mental disorders so severe that they prevent people from being cooperating members of society in the usual sense.
>
> Other questions we can discuss later, and how we answer them may require us to revise answers already reached. This back-and-forth procedure is to be expected. We may think of these other questions as problems of [the] *extension* [of Justice as Fairness].[545]

The idea here seems to be that the groups Rawls mentions, plus others unable to contribute, are to be included within the remit of moral concern, at the very least. It may then be further argued that they are to be included within the remit of justice. I have some things to say about the former claim in subsection 18 below. I argue against the latter claim on two

545 See *PL*, p20, my emphasis. See also Stark (2009) pp. 87—88

grounds. First, Rawls's solutions to the "problems of extension" at best leave it ambiguous whether he can claim that justice extends to the non-contributing. Second, even if this route were to be taken, including the non-contributing on the basis that they partially match up to the ideal of citizens expressed here remains morally problematic. This is not a standard which they should have to meet.

What Rawls explicitly says does not obviously include those unable to contribute on the grounds of disability within the bounds of justice. He states that he cannot see how Justice as Fairness can be extended to the permanently disabled, and that it is likely that it cannot.[546] "It is obvious" he writes "that we have a duty towards all human beings, however severely handicapped."[547] From what I can tell, however, there is no suggestion that this is a duty of justice.[548] These duties could as easily be natural duties given Rawls's system (as proposed by Quong and Kelly).[549] I do not think the natural duties, aside from the duty of justice, are part of justice for Rawls.[550]

Against writers who insist that Rawls includes all human beings within the scope of justice by his endorsement of human rights, I hold that the position of the human rights in Rawls's system is ambiguous. In *The Law of Peoples*, human rights are respected by both liberal democratic societies and decent hierarchical societies.[551] They are a subset of the rights recognised in both kinds of societies.[552] For decent hierarchical regimes, these rights are derived from their "common good idea of justice."[553] A common good conception of justice is obviously different from Justice as Fairness. Even if the human rights might be thought of as a requirement of justice under a common good conception, this does not imply that they are under Justice as Fairness. After all, not all rights in Justice as Fairness are rights of justice. It is perfectly possible that the human rights might be thought to be grounded on justice between the cooperating, and on humanity or mutual aid for the non-cooperating. It is true that Rawls's discussions of human rights in *The Law of Peoples* may suggest that they are minimal rights of justice for all, including the non-cooperating. But he does not make this explicit.[554] Hence we should not simply appeal to human rights as a

---

546 *PL*, p. 21
547 *JF*, p. 176 fn59
548 Stark (2007) p. 130 fn10 agrees Rawls is unclear on this matter.
549 See Quong (2007) pp. 93—97, and Kelly p64
550 See fn 48 above
551 *LP*, pp. 65, 68
552 *LP*, p81
553 *LP*, p65
554 Buchanan (1991) p. 230 fn6 reports a conversation with Rawls. In it, Rawls stated that he believed that those unable to cooperate were owed justice. However, simply him saying this does not mean that such persons can be included in Justice as Fairness given justice as reciprocity. From his subsequent published work, what Rawls reported to Buchanan need not have been his final stance. See, in particular, *PL*, pp. 244—245

quick and easy way to include the non-cooperating within the scope of justice in Justice as Fairness.[555]

It is unclear in any case whether Rawls can include the non-contributing within some minimal provision of justice. None of the authors who suggest this claim also guarantee the non-contributing full liberal rights. But this goes against Rawls's account of the basis of equality. As observed in subsection 5.1, justice is to be granted on the basis of possessing minimal moral psychological characteristics sufficient to fall within the range property of having, or being able to have, the two moral powers. On this basis, how can a person be granted some kind of minimal justice, of the kind that might be thought to be embodied in human rights, but not be granted full liberal justice, if we are operating purely within Justice as Fairness? Once again, it seems that a person's inability to contribute in the right way is being taken as a reason for arbitrary exclusion.

Despite these difficulties, what if we accepted the claim that Rawls only viewed the idea of moral persons cooperating in the maintenance of the basic structure of society as an ideal, and not as a restriction on justice? What this amounts to, however, is to hold that those with the moral powers who are unable to contribute can be included within the bounds of justice solely as non-ideal cases. This is unsatisfactory. Though the members of the well-ordered society can all now recognise that unproductive persons possessing the moral powers can be owed justice, they must still be said to fall short of the basic ideal of the person or citizen. We then have two options. We might hold this basic ideal is a *moral* ideal. If we make the ability to contribute to such arrangements part of a moral ideal, then if people fall short of this in any respect this must be said to be a moral failing. But it is ludicrous to hold that the inability to cooperate is a moral failing. If instead we hold that the ideal is partially a *non-moral* ideal, then we can reprise Barry's criticisms, quoted in subsection 16.4, of the scope of justice being determined on morally arbitrary grounds.

These considerations allow us to elaborate another way in which the contribution requirement is out-of-kilter with Rawls's moral psychology. The alternative ideal of a member of the well-ordered society I am proposing in opposition to the later Rawls is not of a moral person or citizen cooperating throughout a complete life to maintain the basic structure of society. Rather, it is simply of a moral person or citizen. The ideal of such a person is someone with the right attitudes as regards their unavoidable relations that hold between them and the rest of society. It is a person of good will. In order to be a person of

555 In this, I am disagreeing with Freeman (2006) p. 415—416. Similar things to what I have said here could be said about Rawls's distinction between "domestic" or "political" justice, as it applies to the basic structure, and "local" justice, as it applies to the associations of civil society, which are other distinctions Freeman refers to in an attempt to resolve this issue. See *JF* p. 10—12 and *PL* p. 21.

good will, I do not think it needs to be the case that we are able to act as our good will would have us act. What is important is that we are willing to so act if we are able. If such a person is unable to cooperate, they will accept this, and will meet the requirements which can be reasonably asked of them. This acceptance need not be total. They could of course sincerely wish that they were able to cooperate. It is only understandable that this could be a source of regret and sadness in their lives.[556] But ideally we would expect this regret to always be accompanied by self-respect, and pride in their achievements.[557] What this regret shouldn't be associated with are the self-chastising moral emotions of guilt or shame – at least not in the ideal case. For Rawls, these emotions link up to the concepts of the Right and the Good respectively.[558] The ideal of the person engaged in a scheme of cooperation, if a moral ideal, would either be an ideal of the Right or the Good, most likely both. But this would lead to persons feeling guilt, or even more likely shame (as shame is directed towards defects in one's self),[559] if they are unable to cooperate, providing we assume, as we should for the well-ordered society, that they fully psychologically embody the society's ideals. Of course, we can understand that people can feel these emotions over impairments which are in no way their own fault. But Rawls would be unlikely to agree that they *should* feel this way.[560] The contribution requirement, however, appears to commit him to this. Once we eject this requirement, such emotions are, on the face of it, representative of a different set of ideals than Justice as Fairness, or else stem from considering a non-ideal rather than ideal case. Feelings such as these, in an ideal situation, would be irrational.[561]

15.5 B: Isn't dropping the contribution requirement incompatible with the Publicity Condition?

Dropping the contribution requirement does not violate the publicity condition. That condition, it will be recalled (section 2, section 11), states that any adequate

556 *TJ*, pp. 442—443/388, 481—482/421—422
557 See *TJ*, pp. 440—442/386—388 for the self-respect expected in a well-ordered society.
558 *TJ*, p. 482/422
559 The moral emotions literature widely agrees that guilt adheres to one's wrongful, or believed to be wrongful, actions. Shame, by contrast, adheres to the way one is, independently of one's actions. See, for example, Taylor (1985), Wollheim (1999) pp. 155—157. This seems correct to me. If so, it may require a modification of Rawls's account of shame. See Deigh (1983), though note that I do not think that Deigh's account views the matter from a perspective sufficiently internal to Justice as Fairness.
560 More precisely, they may feel natural shame, but they will not feel moral shame. It can be expected that, in an ideal well-ordered society, no one feels natural shame. See *TJ*, pp. 444—445/389—391
561 Roughly, Rawls holds that moral emotions always reflect our genuine moral beliefs (see *TJ*, pp. 481—482/421—422), so the persistence of these emotions in the well-ordered society would entail a conflict in the beliefs of the non-cooperating.

conception of justice must be capable of being publicly shared between the members of a well-ordered society. That one does not hold an employed position in the basic structure of the well-ordered society does not entail that you cannot know the content of the public conception of justice. It also does not mean that others cannot know that you know that conception. Advocacy in some public forum is not required for shared knowledge – we do not need to see our fellow citizens swear allegiance to the state every day, as in the society of Yevgeny Zamyatin's *We*.[562] Rather, common knowledge is had by much more diffuse and indirect means. If I do not need to be personally acquainted with each and every other worker in my country for common knowledge to exist, why think I need to be acquainted with every non-worker?

15.5 C: Isn't it often problematic to find out who is capable of developing the two moral powers? Won't this be even more difficult for those who cannot contribute?

There are problematic cases amongst *both* those who can and those who cannot contribute. That the proportion of problematic cases in the latter category may be larger than the former does not make for a special kind of problem, such that we might think that the non-contributing can be legitimately excluded from justice. Often it will be obvious that a person unable to hold down a normal job or position is nevertheless capable of developing the two moral powers. We can mistakenly think that an individual can develop a sense of justice *both* when they can work *and* when they cannot, when in actual fact they are incapable of developing one. I address such cases in sections 16 and 17 below, but they represent no reason to say that some persons who *are* capable of developing a sense of justice, but for whom it might be difficult to tell whether they can because of various impairments, aren't owed justice.

This again might be thought to pose problems for the publicity requirement. But the equivalent case of the person who is able-bodied, but for whom we are uncertain whether they are capable of the moral powers, is also problematic for publicity. In general, publicity is an ideal to be aimed for. It characterises the well-ordered society, and the well-ordered society itself is an ideal. Such ideals should not be used in order to exclude certain persons from basic recognition simply because their capacity for the two moral powers is difficult to discern (see further subsections 15.5 D and 15.5 E below)

562 *We* by Yevgeny Zamyatin is a Russian dystopia, published in 1921. It was one of the leading inspirations for George Orwell's *1984*.

15.5 D: Isn't it often impossible for society to realise the capacity for the two moral powers in all persons?

Sometimes it is, and this means we are dealing with a non-ideal situation. Non-ideal theory for Rawls deals with two possible deviations from the ideals of justice: injustice arising from people's free choice (either from the active pursuit of injustice, or the passive acceptance of the unjust actions of others), and unavoidable injustice arising from limitations and burdens from one's environment.[563] It is the second kind of non-ideal situation we face when there are persons who have the capacities to develop the moral powers, but whose capacities can in no way be realised. Being in a non-ideal position, these persons are still owed justice: justice which unfortunately cannot be given to them.

In order to get to grips with this topic, we should first distinguish between (1) a person with a capacity for the moral powers, and (2) a human being with the capacity to become a person with a capacity for the moral powers. The may also be a third case: (3) those who possessed the capacity for the moral powers, but have irretrievably lost this capacity. I postpone discussion of this third group of individuals till subsection 18.2.

What actual human beings fit these cases? Intuitively, an immoral person whose immorality stems from a brutal upbringing, but who might be rehabilitated, fits (1). A congenitally psychopathic individual by contrast, fits (2). Such individuals represent just one way in which human beings can fail to possess a sense of justice — roughly through having a lack of empathy.[564] Certain autistic individuals may be unable to develop a sense of justice. But this is a different condition — autistics are not psychopathic. Other individuals with various kinds of brain disorders or brain damage will represent yet more cases. I present this selection of cases in order to emphasise that someone who lacks a sense of justice does not necessarily fit the profile of a criminal psychopath. The diversity of human nature makes things much more complicated than this. I shall return to this issue in sections 17.2.

---

563 See *TJ*, pp. 8—9/7—8, 245—248/216—218, *LP*, p5. I am indebted to Simmons (2010) for clarification on this topic. See esp. pp. 12—18.

564 Note that saying that psychopaths, and those with an upbringing which damages their empathetic capacities, lack empathy (or as it is sometimes called, sympathy) I do not think necessarily tells in favour of morality developing on the basis of psychological reciprocity principles, or psychological principles of altruism, etc. (on these see Appendix II). Hence I do not contradict one of Rawls's basic psychological assumptions. Empathy is usually understood to represent simply the ability to share another's feelings, in certain complicated ways. This is presumably involved in reasonableness as much as in altruism. For the discussion of empathy or sympathy I have primarily drawn on here, which specifically focuses on Hume, see Krause (2008) pp. 79—82. For an introduction to recent discussions on psychopathy, see, for example, Prinz (2007) pp. 42—47

Regarding the two cases we are considering, it must be assumed there is some point – or at least vague expanse, if we take Rawls's point about the vagueness of moral personality seriously (subsection 15.1) – at which the alterations to an individual's nature needed for them to become a moral person are so profound that there no longer exists the relevant relation of personal identity between the prior individual, and the posterior moral person.[565] A *prima facie* case would be the kittens imagined by Michael Tooley in his "Abortion and Infanticide", who *could* be injected with a serum which will lead them to acquire moral personality when they are fully grown cats.[566] The kittens, we assume, were not beings with a capacity to acquire a sense of justice, in the sense Rawls means this. Rather they were beings with the capacity to become beings with the capacity to acquire a sense of justice.

Where the point should be placed in order to divide those with a capacity for a sense of justice, and those with a capacity to acquire that capacity is a problem for all theorists. But providing a solution can be found – one furthermore sufficiently compatible with Rawls's overall position – then we can say that persons with a capacity for the moral powers definitely are owed justice, but human beings with the capacity to become such persons may not be (I need to say "may" for now: sections 16 and 17 further argue that Rawls should hold they are *not* owed justice). The latter individuals are not potential moral persons. They are not persons with a capacity for a sense of justice. Rather, they are numerically distinct[567] from any such moral persons. They are non-moral persons. For such human beings to be transformed into moral persons, such non-moral persons must cease to exist.

Regarding the beings with the capacity for the moral powers, it has undoubtedly been impossible up until now to realise the innate moral nature of each and every one. Perhaps it ever more shall be so. Such impossibilities are either the product of unavoidable burdens on the resources available to society, or else are maintained not through such scarcity but through unjust actions. But in neither case does this mean that such persons should be considered outside the scope of justice.

The following aspects of Rawls's theory might be thought to militate against this conclusion, and that of the previous subsection as well. First, Rawls explicitly tries to frame his theory to fit with the practically possible.[568] The parties in the original position

565 For the notion of the criterion of person identity as applied to persons, see the introduction to Perry (1975)

566 See Tooley (1972), pp. 60—61. The though experiment is meant to cause problems for arguing that potential persons possess rights.

567 For this term, see again Perry (1975)

568 See, for example, *JF*, pp. 2, 185

choose a conception of justice on the basis of human nature. If there are individuals whose sense of justice is always going to be very unlikely to be realised, won't the parties in the original position simply accept that principles of justice should not be framed to include these individuals, on grounds of practicality?

This attitude by the parties is inadmissible, as we are talking about the features of human beings which qualify them to be represented in the original position in the first place.[569] Being representeed in the original position, these individuals will get a veto on which conception of justice is chosen, given the finality condition and the need for all parties to agree. Hence, when a person with a capacity for the moral powers fails to be given what they deserve according to justice, leading to their capacity for the moral powers to fail to be realised (providing this failure is due to not through brute misfortune in favourable circumstances),[570] this is always a non-ideal situation. Ideally, it should be sufficiently likely that in favourable conditions, everyone who has a capacity for the moral powers should have that capacity realised. Conceptions of justice cannot be chosen which would lead to some moral persons *having no possibility* of realising their moral powers, even in ideal circumstances.

A second element of Rawls's theory relates to the above objection. This is that the parties are to choose principles in view of the limitations on information which affect legislation and constitutional design in a liberal democracy, including the burdens of judgement (subsections 4.5, 12.1, and 13.2).[571] Limited information is obviously available about which human beings are capable of possessing a sense of justice, and which are not. However, once again I cannot see how the parties in the original position can take this as a consideration for choosing a conception of justice which systematically excludes such individuals whose capacity for the sense of justice is hard to discern from the scope of justice. For some of the parties in the original position must have representees who are these very individuals, and they will not put aside the fundamental interests of their representees for anything — they will refuse to enter into such a contract.

In summary, the content of the first principles of justice is not susceptible to being altered in order to exclude individuals whose potential for the moral powers is difficult to

569 *CP*, p. 112 is very clear in this regard.

570 Favourable circumstances do not *guarantee* that all moral persons will realise their moral powers. Rather, they allow that public institutions can be set up which give the best chance of avoiding this possibility. This is implied by the two roles of psychology in justification – moral psychology does not guarantee that, in favourable circumstances, each member of the well-ordered society will have their fundamental interests met, but rather shows that a given conception of justice is not certain or near certain to fail to meet those needs for everyone (futility-avoidance), and is comparatively the most stable conception of justice (arbitration).

571 See, for example, *CP*, pp. 346—351, *TJ*, pp. 156/135, 160—161/138—139, 320—325/281—285

discern, or perhaps even impossible to bring about.

15.5 E: Isn't it sometimes difficult to guarantee the basic liberties, and their fair value, to those who possess the moral powers, but are unable to cooperate (usually through certain impairments)?

Again, this is often a problem, and when it is a problem, it means we are dealing with a non-ideal situation. Deep practical difficulties obviously arise in guaranteeing the severely physically or mentally impaired their full basic liberties as they are specified by Rawls. But it has not been sufficiently recognised that these practical difficulties should be considered a product of unavoidable burdens on contemporary liberal society, i.e. non-ideal conditions.

It may be countered here that it is simply unnecessary to give the severely impaired their full basic liberties.[572] I would agree – providing such human beings are so impaired as to lack the capacity for the moral powers. Hence they are amongst the beings described above – those who are so different from moral agents that transforming them into moral agents would require some severe alteration, if not complete change, of personal identity. But, for all who have the capacity, ideally society should be arranged so that they can exercise their full liberal rights.

It has been widely emphasised that the development of a capacity can be encouraged or restricted by the arrangements persons are situated within. Writers on justice and disability widely "reject any assumption of disability as an individual disadvantage, and [present] instead a distinction between impairment, seen as relating to a loss of some aspects of [human] functioning, and disability, defined in terms of the limitations imposed on impaired people by *the design of social structures*."[573] If these are limitations on access to basic justice, they can be imposed or allowed to remain only if environmental burdens or the inertia of existing social arrangements render this the all-things-considered morally preferable option. We would then be in non-ideal circumstances – indeed, in a case in which the general conception of justice, rather than the special conception, is applicable.[574] Outside such burdened environments, institutions providing the basic liberties cannot be arranged simply to be optimally efficient for the wider populace when this requires the exclusion of people with intellectual or physical impairments. The complexity of the state

572 Freeman (2006) p. 415—416

573 Terzi (2010), p. 151. My emphasis. See also Pogge (2010) pp. 30—31

574 On the distinction between the general and special conceptions of Justice as Fairness, and their relevance to the restriction of the basic liberties, see *TJ*, p. 60—63/52—55, 244—248/214—218

apparatus is a means to serving the lives of moral agents. It is not a tool for preventing the development or employment of the moral sensibility of those agents through its complexity.

Throughout this discussion, I have attempted to abstract away from the empirical details as to how many people fit into the different categories I have outlined. But it seems worthwhile here to note that I believe that a fixation on the current arrangements of our basic constitutions and legal systems often leads us to ascribe less competency to the mentally impaired in matters of justice and moral judgement than we should. These systems are as complicated as they are in order to deal with *ourselves* — the complexity of our pursuits and projects, our diversity, and our vices. We – those without mental or physical impairment, or at least those commonly so regarded – rightly judge that many with mental illnesses or impairments cannot be expected to deal with these institutions by themselves. We then, wrongfully, assign the fault to those persons, and assume that we must adopt a wholly paternalistic attitude towards their relationship to the basic institutions of society. I am not saying that, if the structures are to stay as they are, these people do not need help. Everyone needs help in these matters — that is what lawyers, civil servants, nurses and doctors, teachers, university lecturers etc. are for. But the possession of the moral powers only requires that a person understand the basic normative ideals of the society that they live within. I think that the ability to comprehend these ideals is possessed by a great many more mentally impaired individuals than is commonly recognised. Reciprocity, fraternity, freedom, responsibility etc. are not esoteric concepts, however much a full philosophical grasp of them may be. It is having a basic type of stance and set of sentiments towards the others in your society which qualifies you for justice, and it is solely the lack of this endowment which disqualifies. There is no reason to assume that the basic mental and physical preconditions of such a moral sensibility aren't quite minimal.

I have elaborated these points by reference to the provision of the basic liberties, as they must be met most urgently according to Rawls's theory. How Rawls's conception of distributive justice, as governed by the second principle of justice, may have to be reformed, I do not engage with here. It may be that Rawls's formulation of the difference principle needs to be changed.[575] For persons unable to cooperate, the equal opportunities principle appears, on the face of it, inapplicable.[576] No matter what revisions are necessary, I believe the contribution requirement must be rejected. It is fundamentally inconsistent with Rawls's basic moral and psychological assumptions, and his account of the basis of equality.

---

575 For proposals on how to avoid this, which may or may not be compatible with my position, see Stark (2007) pp. 136—140 and Richardson (2006) pp. 430—439

576 See Stark (2007) p. 134 fn22

15.5 F: Isn't this revision incompatible with Rawls's resourcism?

Rawls's resourcism comes from his use of the social primary goods as a metric of justice, and the fact that the distribution of these primary goods must be seen to be publicly ascertainable. It might be wondered whether distributing to everyone possessing or capable of the moral powers – including those severely impaired – could be accounted for. I have already argued that the non-contributing do not pose a special case as regards the various aspects of publicity (15.5 B and C above). The position of being a non-contributing-but-moral agent, and the resources distributed to you in light of this, is as capable of being publicly verified as anything else.

It is important to note that the issue I am concerned with can be considered separately from the debate surrounding whether primary goods are the correct metric of justice. Rawls pioneered the primary goods or resource metric. In this, he is joined by many others: both broadly Rawlsian in their approach and not.[577] What is to be distributed are simply valuable resources, including public institutional arrangements as well as exchangeable commodities.[578] A leading rival is the capabilities metric, as proposed by Amartya Sen and Martha Nussbaum.[579] For them, it is capabilities that are to be distributed justly. Capabilities are relations between external resources and the internal dispositions of persons to realise valuable "functionings": things persons value doing or being.[580] Resourcist positions are blind to variations in capabilities — at least, beyond the capacities to develop some basic moral sensibility, such as Rawls's two moral powers.[581]

The chief claim of the capabilities theorists is that primary goods do not represent a fair metric, as different people possess different abilities to convert the same resources into valuable functionings. One chief group who are said to come off unfairly under primary goods are those with severe impairments – as we have noted, a group which overlaps significantly with those unable to contribute to the basic structure of society. In response, resourcists have claimed that the disabled need not be said to be unfairly treated under a primary goods metric, if the implications of such a metric are properly understood.[582] But

577 For examples of the former, see Pogge (2010) and Kelly (2010). For examples of the latter, see Dworkin (1981) and Carter (2011)
578 See *CP*, pp. 271—273
579 See, for example, Sen (1980) and Nussbaum (2006)
580 Many of the other prominent metrics are varieties of welfarism, and take the realisation of well—being to be what we should attempt to distribute justly. See, for example, Cohen (1989).
581 See *PL*, pp. 182—183 and *JF*, pp. 169—170
582 See, for example, arguments by Pogge (2010) pp. 32—53 and Kelly (2010) pp. 66—69

all I have claimed is that anyone with the capacity for the moral powers is owed full justice, regardless of their ability to contribute. All theorists should agree that people with this basic moral sensibility, or the equivalent found in their own theory, are owed equal recognition.[583] Some have argued that only a resourcist position is compatible with this basis of equality.[584] Others have argued that the capabilities approach is best compatible.[585] Whichever is the correct conclusion I can leave aside here.

15.5 G: Isn't this position incompatible with the basic structure being the first subject of justice?

It might be thought that by jettisoning the contribution requirement, I am hence committed to siding with the recent critique of Rawls's distinction between principles of justice for the basic structure, and for individuals. This criticism has been proposed by G.A. Cohen and Liam B. Murphy, and has been accepted by Rawlsian philosophers such as Michael Titelbaum.[586] These philosophers argue that there do not exist principles of distinct content governing the arrangements of the basic structure of society, and the actions of individual agents within the rules of that structure, in opposition to Rawls.[587] Others have produced various counterarguments for the correctness of the distinction.[588]

My revision holds independently of the basic structure debate. Whether or not one thinks distinct principles exist for the basic structure, everyone agrees that the members of the well-ordered society share a public conception of justice, which embodies certain fundamental normative ideas. My observation is that society as fair cooperation, and justice as reciprocity, impart normative ideals into Justice as Fairness which are at odds with aspects of its moral psychology, and in addition are independently morally dubious. Cohen and Titelbaum also agree that aspects of Rawls's psychology undermine the basic structure restriction.[589] But, though these two criticisms might stem from the same source, they may be sustained separately.

It may be possible that someone could argue for the following interpretation of the well-ordered society. Those unable to contribute to society are fully included within the scope of justice. But everyone, including the non-contributing, understands the principles

583 Sen (2010) pp. 242—243 is clear that he does not disagree with Rawls over this issue.
584 See Carter (2011) pp. 560—571
585 See Anderson (2010)
586 See Cohen (2008) chapter 3, Murphy (1999), Titelbaum (2008)
587 Relevant passages include *TJ*, pp. 47/54, 108—110/93—95, 116—117/99
588 See, for example, Williams (1998)
589 See Cohen (2008) pp. 129—132, Titelbaum (2008) pp. 296—302

of justice as being quite distinct for institutions and individuals. Institutions need to guarantee justice for all, but, outside the rules of the basic structure, everyone, again including the non-contributing, commit no injustice when they fail to observe individualistic correlates[590] of the principles for institutions in their own choices. I have sympathies with those who reject the distinction between institutional and individual principles, and it is obvious that the considerations I have stressed are similar to the ones that these other authors stress. Both criticisms proceed, in part, from considering the fundamental attitudes and perspectives that the members of the well-ordered society can be expected to have towards one another on alternative conceptions of justice. But I leave open the possibility that one criticism might fail while the other might succeed.

15.5 H: Couldn't simply expressing the moral powers in any sense be said to be a "contribution" to society for Rawls?

If this were true, my criticism of Rawls would lose some of its bite. For then it would be the case that to possess the moral powers would be to possess the ability to cooperate after all. I earlier rejected the suggestion that to have the ability to cooperate in the maintenance of the basic structure through normal employment is what amounts to having the moral powers. This suggestion is that even the most generic activities of being a moral agent could count as contributing to the cooperative scheme of society with one's fellows, in which goods are produced which benefit all fairly.[591]

It should be clear from how Rawls defines cooperation in the basic structure that this is not the case for him. Productive activity is assumed to be on a wider scale than the most basic moral agency. Persons can be moral agents, and yet be unable to contribute to the minimal maintenance of their society from one generation from the next. A society in which each member was afflicted by a sufficiently incapacitating physical disability would be rendered unable to feed itself, and hence, despite the moral sensibility of its inhabitants, would perish. Consider, for example, the scenario described in John Wyndam's *Day of the Triffids*, in which nearly the whole of the population lose their sight, and are hence left defenceless against the carnivorous Triffid plants.[592] The nature of the disaster in the book could have easily left everyone blind, and hence could have led to complete destruction.

There is something weak about this point. It is the case that many of the

590 That the principles of justice would be correlates to Rawls's principles for institutions, not the same principles themselves, see Titelbaum (2008) pp. 293, 302—307

591 This kind of defence of contract theory is suggested by Silvers and Francis (2005) and Hartley (2009)

592 John Wyndam was a British science fiction author. *The Day of the Triffids* was published in 1951.

contributors in a well-ordered society perform roles which are not strictly speaking essential for the maintenance of the basic institutional structure in the society year after year. And wouldn't being a morally engaged individual in society help to support the public culture of the society, even if one could not hold down essential paid employment? I am unsure what to say here. Rawls's ideal of a member of the well-ordered society is an individual who works, and acts so as to at least not undermine the public political culture of the society (see the description of the duty of justice in section 2).[593] I do not believe that my revision can be incorporated into Rawls's theory without some degree of substantial revision. But the precise extent I shall not investigate here.

However, we can accept that Rawls would recognise that, though society could not be sustained on the activities of the non-contributing alone, there could be many ways in which these people could contribute to the maintenance of aspects of the basic structure, either directly (i.e. speaking or communicating on public forums) or indirectly (i.e. through helping to support sentiments of friendship, trust, and fraternity).[594] Even with this, however, my point would not be otiose. For it is not in virtue of being a contributor in any kind of way that people should be valued, but as moral agents with good will towards their fellows. That the activity of being such an agent produces some kind of benefit is beside the point. On occasion, good intentions may lead to bad outcomes. No matter how many times someone's good will went awry, however, we would not be warranted in excluding them from justice (though this is not to say they should avoid all sanction). Reciprocity, I argue, at bottom implies the mutual concern for our own and each other's good, and *hence from that concern*, the production of mutual benefit, ideally.

15.5 I: Isn't this revision ruled out by Rawls's conception of reasonableness?

It will be recalled that reasonable persons desire for its own sake a social world in which they fairly cooperate with others for mutual benefit (subsection 8.2). To be a reasonable person, it is not required that you are able to fairly cooperate with others. It is enough that you desire to. This is clear from Rawls's claim that being a reasonable person binds *in foro interno*.

593 I put this quite weakly as, as was noted in chapter 4, Rawls cannot demand that the members of a well-ordered society be necessarily strongly politically engaged.

594 On the possibility of the latter, see Silvers and Francis (2005)

15.5 J: How does this revision impact on Justice as Fairness as a political conception? Is the revision compatible with political liberalism?

What is distinctive of political liberalism (section 13), as (perhaps) opposed to public justification liberalisms generally, is that it demands that political conceptions be presented as only reasonable and not as true. This does not significantly alter the normative content of Justice as Fairness, but only regarding the way public political arguments are to be presented. The account of the scope of justice appears to be untouched. Hence I believe this revision is perfectly compatible with political liberalism – simply that it might be more demanding or controversial does not, in itself, debar it from being part of public reason. Who is to say it may not actually be better founded given certain public political cultures?

15.5 K: Does this revision lead to any alterations in Rawls's fundamental ideas?

Yes. The alteration stretches to a rearrangement and a slight change in content. It was remarked above in subsection 15.1 that society as fair cooperation is the most fundamental idea in Rawls's theory, and that the idea of the person and the idea of the well-ordered society are defined by reference to it. This cannot be sustained if we are to drop the ideal of the person as a contributor to the maintenance of society from one generation to the next. Instead, the idea of the person as possessing the two moral powers must be thought of as the most fundamental. It can then be joined with the idea of the well-ordered society to yield the modified idea of society as fair cooperation. That society can then persist from generation to generation is then added as an empirical postulate, not a normative ideal (or at least, not a moral one, by reference to which individuals or societies can be morally judged). It is, of course, non-ideal (to say the least) if a well-ordered society were to collapse due to mass blindness and triffid attack, as in my example above. But the collapse here is due to environmental burdens. The situation is not non-ideal in a moral sense if the members of such a society act in a morally peerless fashion, despite the hopelessness of their situation.

There is a further aspect of Rawls's philosophy which may need to be revised in the light of this alteration. But the matter is difficult to discern, and would take some work to fully analyse, due to the difficulty in working out just what Rawls is committing himself to. The aspect I am thinking of stems from the principal sense in which Rawls takes his theory to be distinct from Kant's. Kant, Rawls claims, takes as the basic unit of morality the

individual moral person. Rawls by contrast, conceives of morality, or at least justice, in a more fundamentally social way. He takes the basic ideas to be the moral agent within a moral society – the person within a well-ordered society, in other words.[595] I am unsure whether this assumption can be maintained in the face of the revision I am proposing. It may be that it can. For consider, the members of society who are impaired in such a way to prevent them from contributing to the maintenance of their society can still conceive of the ideal as their being able to contribute. Of course, this must be a non-moral ideal, for reasons outlined in 15.5 A above. I can see potential problems with such a response, but I do not pursue them here.

15.5 L: What about non-contributing groups other than the mentally or physically impaired?

The two other major groups of non-contributors are certain animals, and future people. Some animals have contributed to society for centuries — they are excluded on the assumption that they cannot possess the moral powers. I leave aside these animals, and the animals undoubtedly unable to contribute, until sections 16.2 and 18.

Future persons are unable to cooperate for mutual benefit with the people of the present in terms of material goods. It is less clear whether they can be said to cooperate when they keep alive valuable practices and institutions. In any case, persons are owed justice on the basis of their moral powers. Either future persons will have such powers or they will not. So long as they do, we owe them justice – their temporal displacement from us being morally irrelevant.[596]

15.5 M: Does dropping the contribution requirement lead to any alterations in the original position?

Yes. Those who are unable to contribute to the maintenance of the basic structure must now be included within the original position. Under the veil of ignorance, no one will know whether they will be able to cooperate in the maintenance of society or not. As I indicated in subsection 15.5 E, this may lead to alterations in the principles agreed to in

595 See, for example, *TJ*, pp. 256—257/226, *CP*, p. 340

596 This position differs from Rawls's own account of how we owe the people of the future justice (*TJ*, pp. 128—129/111, 139—140/120—121, *PL*, p. 274) I do not explore whether this demands alterations to Rawls's account of a just savings rate between generations (*TJ*, pp. 284—293/251—258)

that position. Other writers have suggested similar modifications.[597] I believe the alterations entailed for Rawls's theory go further than these writers indicate. I do not pursue this matter here.

15.5 N: Doesn't this alteration to Rawls's theory disrupt his account of international justice?

My arguments against the contribution requirement may have left some wondering about the status of persons who wish to become citizens and contribute to different societies. For, if persons born into a certain society are to be granted full justice and equal liberal rights simply because they *wish* and *are motivated to* contribute to a society, even if they can't, what is to stop a foreign national claiming citizenship of that society simply because they conscientiously wish to contribute to it?

I believe that Rawls's theory allows us to say the following, and hence that this proposal does not constitute a serious *internal* problem for a Rawlsian theory.[598] For Rawls, being a just person requires you to attempt to promote and sustain just institutions, so long as the (moral) costs are not too great.[599] There is no restriction to the institutions of your own society, though of course the task of promoting justice abroad is complicated by various additional moral-philosophical and practical matters.[600] However, the question of migration from one society to another is another matter. The duty to promote just institutions does not require an open-borders policy with regards to anyone who wishes to come and work in your society.[601] The duties to (and expectations from) those born within one's society who possess the capacity for the moral powers, but are unable to contribute to the maintenance of society's institutions, are different to those outside one's society.

15.5 O: Justice as Reciprocity or Impartiality?

Finally, over this subsection and the next, we can now return to Barry's question regarding the coherence of justice as reciprocity as a position. Given what has been argued, it does not seem that justice as reciprocity as Rawls understood it is really a stable position.

---

597 Stark (2007), Richardson (2006)
598 Whether an external problem remains, due to problems with Rawls's position on global justice in general, I leave aside.
599 *TJ*, pp. 115/99, 334/293—294
600 See *LP*, pp. 37, 105—113
601 *LP*, pp. 8—9, 39 fn48

Deep-seated elements in his theory undermine it, and there does not seem to be any way to ground it in a morally principled manner.

However, this does not necessarily undermine the acceptability of the slogan. For there is a sense in which, even with the modification I have proposed, a basic idea of justice still remains one of reciprocity. This can be seen by considering Rawls's moral psychology. In acting justly towards persons as they develop a sense of justice, and an ability to develop their own conception of the good, what we expect is that the person develop those capacities to the best of their ability. If they possess the capacity for the moral powers, then, ideally, our expectations will be met. The relationship between persons in the well-ordered society is founded on reciprocity of basic recognition and good will.[602] Reciprocal production of a certain set of goods (i.e. the institutions of the well-ordered society) can be an aspect of this reciprocity, but is not an essential aspect. Non-ideal, regrettable conditions can undermine it.

Does this serve to render justice as reciprocity equivalent to justice as impartiality? I do not think so. It appears that Barry's basic criterion for being included within the scope of justice is to have interests. Hence, animals are included within the scope of justice, though they are only afforded minimal and not full justice.[603] I have assumed the position that we cannot grant a being partial justice (subsection 15.5 A) – beings deserve full justice, or they do not. I hence stick with justice as reciprocity – for now (see section 18 below).

15.5 P: The circumstances of justice, and justice as reciprocity

Rawls's commitment to the circumstances of justice has been argued to be the basis for his commitment to the ideas of society as fair cooperation and justice as reciprocity, with their implication that non-contributors are excluded from the scope of justice. Brian Barry argues the following: Rawls's acceptance of the circumstances of justice, and in particular the assumption that social cooperation, and hence justice, supposes rough equality, leads to Rawls distinguishing – sometimes implicitly, sometimes not – between contributors and non-contributors.[604] This distinction plays itself out in different ways for different characteristic groups.[605] The distinction is traced back to a tension between the idea of impartiality embodied in the original position, which requires us to abstract from

602 A similar reinterpretation of Rawls is suggested by Ci (2006) pp. 74—92, 136—141, 146—152
603 See the page references in fn19 above.
604 See Barry (1989), pp. 179—183, 241
605 See in particular, ibid. pp. 183—189 (on future generations), pp. 203—212 (on animals), pp. 183—189 (on international justice).

contributive capacity, and Barry's interpretation of the circumstances of justice, which leads to justice exclusively being a virtue of schemes between cooperators.

My earlier discussion of the circumstances of justice in section 7 should indicate the problem here. The circumstances of justice merely serve to make social cooperation possible and necessary. They indirectly give rise to the need for justice, but it is *never* explicitly stated that they restrict the scope of justice.

Section 16: The ability to cooperate as sufficient

16.1 Further sufficient grounds?

The previous section focused on defending one idea – that the capacity for the moral powers is sufficient to be owed justice. By implication, the capacity to contribute to production through the basic structure of society is not part of a joint necessity with the moral powers. I also believe, from a Rawls-esque contractarian standpoint, that the capacity for the moral powers *is* necessary to be owed justice. This is contradictory to what Rawls allows in *A Theory of Justice*. There, Rawls holds that being able to develop the two moral powers is only sufficient, and not necessary, to be included within the scope of the rights and duties of justice.[606]

Just which further groups, in addition to those beings with the capacity for the moral powers, can be owed justice is unspecified there. I shall consider three relevant groups of individuals who could be thought to be owed justice despite not having the two moral powers. There are those who have a capacity to develop and revise a conception of the good, and are able to cooperate in the fair scheme of society, but lack a sense of justice. There are those who lack the ability to cooperate, and lack a sense of justice, but can form a conception of the good. And then there are those who lack both moral powers altogether and, we can assume, any cooperative ability. I shall consider the first group in this section. I shall reject the thought that they are owed justice on the basis of their ability to cooperate – the ability to cooperate is not sufficient to be included within the scope of justice. My analysis in this section (section 16) will indicate why I feel I can consider this group, and the second and third groups all together in section 17. In section 17, I shall reject the idea that any of these three groups are owed justice.

As before, first I will lay out independent intuitive grounds that count against

606 *TJ*, p. 506/442—443

including beings without a sense of justice but possessing an ability to cooperate, presenting an external critique. Then, I indicate how elements of Rawls's own theory support this conclusion, thus presenting an internal critique. As before, many of the elements of the internal critique stem from Rawls's moral psychology. This all occurs within subsection 16.2. I then reply, in subsection 16.3, to a recent argument that contributing to a cooperative scheme has its own moral standing. Relying on earlier material, I answer that this proposal cannot be maintained, at least by Rawls.

16.2 Excluding the irredeemably unjust but cooperative

I first present independent moral considerations against the capacity to cooperate being sufficient to be owed justice. I begin by reiterating how it is possible that men and women might cooperate in the maintenance of a just society, despite lacking even the capacity for a sense of justice. Under certain circumstances, individuals without a sense of justice can be moved to act justly so long as it is in their own interest (subsection 15.3). It cannot be assumed that these circumstances will obtain very often within our societies. As I noted earlier in the discussion of the circumstances of justice, in our world, rough equality does not obtain between all agents (subsection 7). So long as they can avoid the reprisals or restraint of the just, nothing will stop those without a sense of justice mistreating those who are weaker than themselves. The irredeemably unjust will always be willing to act contrary to justice when they think they can get away with it. However, it is perfectly possible that overall society may be a net beneficiary from the periods when these individuals do rationally curb their short-term self-interest for the sake of long-term gains, and hence contribute to the overall stock of goods. Expanding on these facts will allow us to develop the best picture of the nature of our relations to those who lack a sense of justice.

Lacking a sense of justice, our relations towards these individuals can be best viewed as equivalent to our relations to animals (or at least most animals, depending on one's views), or perhaps even forces of nature. We can benefit overall from our interactions with nature. But we do not see ourselves as bound to give the same sort of respect to nature as we do to other moral agents (see further section 18 below). Imagine a village where people are occasionally attacked by an escaped panther. The village might overall benefit from the cat's presence in some way. Cryptozoologists may come and spend lots of money staying at the village pub. Merchandising opportunities might exist: mugs and baseball caps might be sold to both visiting tourists, and over the Web. If the choice was offered between keeping things as they are, or else changing matters so as to prevent the beast's

attacks, but only at the cost of also losing the beast-benefits, it would constitute an injustice, all-things-considered, for the leaders of the village to keep things as they are. For this would essentially be to trade off individuals' basic rights against wider benefits for the community, no matter how great those benefits were. There could be, of course, a range of possibilities with regards to how things could be changed. It might be there was only one option: to kill the beast. But this is unlikely, at least in a modern setting. Some kind of protection against the beast, such as a fence, or a beast-tracking system, could instead be used. Some solutions may allow the eradication of the danger from the beast, but the preservation of the benefits it brings. But the important point is that the basic rights of the individual members of the village must take priority.

The villagers' position as regards the beast is analogous to our position regarding humans who lack a sense of justice but who are overall net benefactors to society. We know that they cannot be relied on to avoid injustice. We must take the necessary measures to avoid this injustice. It will be acceptable for us, it might be thought, to enter into productive relations with these individuals if this is compatible with preventing them from committing injustice. But if it is not, then such relations are out of bounds. Failing to recognise our proper relations to such individuals, then, can easily lead to us acting unjustly towards each other. In failing to acknowledge the distinct status of those who are unable to be just, we place each other in harm's way of these individuals.

None of this is to say that we are allowed to do whatever we like to those without a sense of justice, any more than that we can do whatever we like to the animals. Rawls holds that "the capacity for feelings of pleasure and pain and for the forms of life of which animals are capable" impose on us "duties of compassion and humanity." However, they do not impose duties, or expectations, of justice.[607] Our duties towards the animals may very well be much more expansive than are usually recognised. Nor need we assume that humans who lack any capacity for the sense of justice could only be afforded the rights of animals. My discussion so far has been misleading in the following way: it may have put the reader purely in mind of the dangerous psychopath. But not all human beings who lack the capacity for a sense of justice can be assumed to be dangerous psychopaths. My basic point here should be taken to be: whatever we owe to those without the capacity for a sense of justice, it is different to what is owed between persons with the capacity for a sense of justice.

Finally we might point out the implication of including within the scope of justice those incapable of developing a sense of justice but with the capacity to contribute, but not

607 See *TJ*, p. 512/448

those capable of a sense of justice but who cannot contribute. If we do this, then it seems obvious that we are actually valuing contributive potential *more* than possession of a sense of justice. For we are willing to include someone on the basis of contribution alone, but not on the basis of having the capacity for a sense of justice alone. Even if we instead adopt the idea that it is merely a moral ideal that persons be able to take employment in the basic structure, rather than a strict restriction of the scope of justice (as canvassed in subsection 15.5 A above), it then appears that we value contribution *equally* to possessing a sense of justice. For we are willing to grant basic justice on the basis of either contributive potential, *or* possession of the capacity for the moral powers. This is intuitively the wrong attitude for the Just to have.

Having presented external considerations for revising Rawls's theory, I now present internal ones. Several aspects of Rawls's theory militate against including those lacking a sense of justice within the scope of justice, even if they are able to contribute to society. First, Rawls holds that those without a sense of justice lack the moral emotions. Rawls holds that one can only experience the moral emotions if one is able to grasp moral concepts. Hence, justice is beyond the ken of those without a sense of justice. Without such background beliefs, one can experience only anger and regret, not true resentment, indignation, or guilt.[608] In "The Sense of Justice", Rawls claims that a propensity for the moral sentiments and emotions is needed if one is to be able to complain of injustice, as "the duty of justice is owed only to those who can complain of not being justly treated."[609] This rests on the more basic idea that "if a person has a right to something, it must be that he can claim it and protest its not being given him."[610] Rawls dropped this later claim, presumably along with the claim that a sense of justice is necessary for justice. But it does follow from assumptions about the moral emotions which are retained in *Theory*. Unless his arguments for the moral powers being merely sufficient are sound (see section 17.2), this commitment should be maintained.

Furthermore, those without a sense of justice seem incapable of realising the ideal of reciprocity that is embedded in Rawls's principles of moral psychology, as we have described that ideal (subsection 15.4). The essence of their condition is that they are unable to recognise others as possessing the moral powers. They are different, then, from those who recognise these powers but who are unwilling to act accordingly. They are also

608 *TJ*, pp. 487—490/427—429. See also *CP*, pp. 111—112.
609 *CP*, p. 114
610 *CP*, p. 114. This statement appears to assume a choice theory of rights. Some believe there are problems with including children in a choice theory in the way that Rawls proposes to in *CP*, p. 114 and *TJ*, p. 509/445—446. See, for example, Steiner (1994) pp. 245—246. I cannot enter this debate here. I simply assume that those with a capacity for the moral powers can somehow be shown to be owed rights.

different from those who recognise those powers, and the reasons they give, but find themselves unable to act appropriately, i.e. who suffer from weakness of will over this matter. For both these latter groups possess a sense of justice. They will be able to comprehend that their actions or lack of actions are unjust, and hence mandate some kind of response from the just. Both groups recognise that justice requires something of them. This is psychologically impossible for the first group. If you are unable to recognise the sense of justice in others, other things equal, it follows that you lack a sense of justice yourself.

It is true that those without a sense of justice can enter into certain reciprocal relations with the just, under certain conditions. But not any old reciprocation, or ability to reciprocate, can give rise to claims of justice. This is obvious when we consider that reciprocal relations can exist between persons which constitute an injustice to others. As the discussion of the circumstances of justice should have made clear (section 7) not all social cooperation is fair social cooperation (see also Appendix II). Indefensible exclusionary social practices can be perfectly reciprocal between those who are included – honour amongst brigands is still honour, even if restricted. What is required is the right kind of basic relationship of reciprocity: one which serves as a proper basis of equality between all moral agents.

Now it may seem that I am unfairly pressing my case here, by allowing the easy association between those without a sense of justice, and egoists, to pass without comment. For Rawls, these groups coincide. He argues that, all-things-considered, the development of natural attachments and ties to particular others is accompanied by or leads to the development of the moral sentiments, including the sense of justice.[611] This leads him to conclude that those who possess any attachments to others at all are at least capable of developing a sense of justice, and that those who lack the capacity for a sense of justice also lack the capacity to care about others.[612]

Even if we allow, contra-Rawls, the possibility that people can lack the capacity for the sense of justice, but still be able to form ties with particular others, however, I maintain that they cannot be owed justice. The relationship between these individuals, and those outside their circle of relationships, is no different from that between the egoist and other persons in general. Those with a sense of justice who do share attachments and sentimental ties with this individual will be allowing the possibility of injustice if they grant them the privileges and responsibilities of the just. This is not to say that no moral requirements then

611 See *TJ*, pp. 485—487/425—427
612 *TJ*, pp. 487—490/427—429

obtain. If you are friends with a person who is incapable of justice, if you abandon your friendship due to your friend's lack of justice, then you violate the ideals of friendship. But, so long as the requirements of justice and friendship can come apart, you must act as a friend to your friend, but act justly towards the just. For better or worse, these two values can compete.[613]

Finally, it may be wondered whether Rawls truly thought that justice could be extended to those who lack a sense of justice but who are able to cooperate. I believe that the jury is out, due to the brevity of Rawls's discussion of the moral powers only being sufficient. Some have interpreted him to be talking about only the mentally and physically impaired in this section.[614] Nothing I see suggests this, however. Whatever he meant to suggest, it seems that including the cooperative but unjust is incompatible with his account of the moral psychology of Justice as Fairness.

16.3 The moral status of contribution

In a recent article, Cynthia A. Stark has defended the moral salience of both of the elements of the ideal of Rawlsian citizenship we identified above: the capacity for a sense of justice, and the ability to contribute. The moral relevance of contribution is based on what Stark calls an anti-exploitation principle:

> The contractarian's commitment to the anti-exploitation principle ... requires him to treat the ability to cooperate as morally relevant because the anti-exploitation principle regards the fact of one's social cooperation as bearing upon what share of the social product one is owed; a person contributing to a scheme of cooperation is owed, *due to her contributing*, a certain portion of the cooperative surplus (all things being equal).[615]

This principle, and a principle expressing the idea that moral persons are owed equal recognition and respect as such, are said to be separate and distinct parts of our pre-theoretical notion of justice.[616] If we neglect the moral relevance of the capacity for the

613 In *Theory*, Rawls has arguments which may suggest that in a well-ordered society, the demands of our personal attachments and the demands of justice and our wider society, cannot come apart (see esp. pp. 474—475/415—416). In later work, however, Rawls appears to be much less sure of this claim. See, for example, *PL*, pp. 57, 197—198.
614 For example, Nussbaum (2006) and Freeman (2006)
615 Stark (2009) p. 90
616 Ibid.

moral powers, then our theory may be impossible to justify to the non-contributing. If we neglect the moral relevance of contribution to the cooperative surplus of society, then our theory may be impossible to justify to the contributing.[617]

I have no reason to deny that, pre-theoretically, we may view justice as having these distinct elements. But, if we accept Rawls's account of the basis of equality, the capacity for the moral powers must be viewed as more fundamental. It is the basis from which the other element of justice is derived (subsection 15.1). We expect things from moral persons, and if they fulfil these expectations, we hold that there are certain things they should receive. When people meet expectations that they should *work*, and they do not receive what they should in return for this, we call this exploitation. Work, and exploitation, are just two of the moral concepts associated with the relations moral agents should maintain between each other. Equal recognition as moral agents is at the basis of both these concepts, as well as others.

Contribution to a cooperative surplus does not have independent moral standing. If an individual who lacks the capacity for a sense of justice has contributed to a cooperative surplus, this does not mean they are owed remuneration in anything more than a conventional or legal sense. The reason for us interacting with the unjust is that we are able to benefit from them without allowing an injustice to befall any of us. If we cannot guarantee our security, there cannot be a reason for us to interact with them. We give them things in return for what they do in order to prevent them from causing injustices, not because they are *owed* them for what they have done. As I have noted, we can owe them other duties. But we cannot owe them duties of reciprocity. We cannot exploit them, in the relevant sense. We can only exploit them in the sense that people speak of human beings "exploiting" nature. This is a corruption of the precise meaning of exploitation – it simply refers to misuse. As noted above, they cannot complain of our having this attitude towards them, as they cannot recognise that we ourselves are owed this attitude.

Stark identifies as equally fundamental two principles of justice which in fact occupy different levels of the concept. Exploitation is relevant between those who contribute to the cooperative surplus of society. But it presupposes a more basic, and expansive, conception of the duties and rights of a moral agent. Justification between persons begins from this basic equal status, and then moves outwards through each of the various social positions and roles that persons are able to occupy.

The arguments I have presented here and in the previous subsection cast doubt, at least, on the claim that the ability to contribute to the cooperative scheme of a well-ordered

617 Ibid. p. 91

society is sufficient, in itself, to be owed justice. It now remains to ask whether those without the capacity for the moral powers are to be included within the scope of justice on the basis of some further attribute.

Section 17: Those who lack moral powers and the ability to contribute

17.1 Those without a sense of justice aren't owed justice

I have argued that possession of the ability to contribute to the maintenance of the basic structure of society is, by itself, insufficient to be owed justice. Such individuals who possess this ability, but lack the moral power of the sense of justice, are in the same boat, morally, as the remaining two groups which I shall consider. These are those who lack a sense of justice and an ability to cooperate, and those who lack any cooperative ability or moral powers. I shall assume that the bare minimum to be included in these three groups is simply that one have interests, and shall from now on treat these three groups as a single group. I hence assume that having interests is not equivalent to having a conception of the good, as this seems plausibly tied to the notion of rationality, and hence freedom (section 8). Nothing essential turns on this assumption.

Is having interests sufficient to be owed justice? Rawls should claim not. I shall only add a little to what has already been said in the previous section here. In subsection 17.2, I shall then address Rawls's own arguments that possession of the moral powers is only sufficient, and not necessary, to be owed justice.

The considerations put forward in the previous section already take us half way to our eventual conclusion. Justice cannot be given to those without a sense of justice. Giving them the rights and privileges of justice, and expecting them to act on the duties of justice, will, in most circumstances, put those who possess a sense of justice at risk. This is in effect to commit an injustice towards others who have a sense of justice, by leaving them open to be preyed on by the unjust. However, some of the group we are now considering can be assumed to be incapable of causing injustices to occur, if their impairments are severe enough. Nevertheless, we incorrectly grasp the scope of justice when we then conclude that these individuals can be granted justice due to the lack of threat they pose. For it is undeniable that, if these individuals were capable of injuring us, there is nothing in their character which would prevent them doing so. In a similar way that we can fail to acknowledge the good will of those with a sense of justice who are unable to cooperate,

similarly we can fail to acknowledge the lack of good will in those who lack a sense of justice but who are unable to harm. If we view these two groups as both being owed the same, on the basis that both have interests, we fail to acknowledge the motivation to reciprocate found in the former group. In essence, we fail to give the willingness to reciprocate its proper due. Assuming that duties and rights of justice are founded on reciprocity, then justice cannot be extended to those who are unable to respond in kind.

17.2 Rawls's arguments for the sufficiency of the moral powers

As noted earlier, Rawls backs away from his earlier stance that the moral powers are necessary and sufficient to be owed justice, and in *A Theory of Justice* believes them to be merely sufficient. What are his arguments for this change?

Below, I quote the whole relevant passage

> The capacity for moral personality is a sufficient condition for being entitled to equal justice. ... Whether moral personality is also a necessary condition I shall leave aside. I assume that the capacity for a sense of justice is possessed by the overwhelming majority of mankind, and therefore this question does not raise a serious practical problem. That moral personality suffices to make one a subject of claims is the essential thing. We cannot go far wrong in supposing that the sufficient condition is always satisfied. Even if the capacity were necessary, it would be unwise in practice to withhold justice on this ground. The risk to just institutions would be too great.
>
> It should be stressed that the sufficient condition for equal justice, the capacity for moral personality, is not at all stringent. When someone lacks the requisite potentiality either from birth or accident, this is regarded as a defect or deprivation. There is no race or recognised group of human beings that lacks this attribute. Only scattered individuals are without this capacity, for its realisation to some minimum degree, and the failure to realise it is the consequence of unjust or impoverished circumstances, or fortuitous contingencies.[618]

618 *TJ*, p. 506/442—443

Later he adds

> It is reasonable to say that those who could take part in the initial agreement, were it not for fortuitous circumstances, are assured equal justice.[619]

It will be clear from the first paragraph that Rawls gives no explicit argument for including those without the capacity for the sense of justice within the scope of justice. Instead, he is simply going to assume that just about everyone has this capacity. Because of this assumption, there is no need, or at least less need, to argue as to whether the sense of justice is merely sufficient. Even if it is necessary, it will be in practice be unwise not to act justly towards any human being.

This last practical suggestion I can partially agree with. My discussions in the previous section assumed perfect knowledge of who is capable of a sense of justice and who is not. But faced with imperfect knowledge, we will be better to give individuals the benefit of the doubt, rather than not afford them justice. Providing that Rawls's empirical assumption is correct, we do not seemingly place those capable of justice in danger by generically extending justice to everyone. Indeed, we may place those capable of justice in danger by *not* so extending matters. However, I do not think such an extension of justice is best viewed as extending the scope of justice in a *moral* sense. Rather, within an overall Rawlsian position, I believe that such an extension is best thought of in a *legal* or *conventional* sense, in order to secure better prospects for protecting the rights of individuals who fall under the scope of justice in a moral sense.[620]

In addition, however, it is not obvious that we will always have imperfect knowledge of who is included within the moral scope of justice. We *do* know of certain individuals who are verified criminal psychopaths. Obviously, we section such persons to protect the rest of society. Intuitively, I am unsure as to whether this is a matter of justice. From Rawls's presuppositions, it is not. Furthermore, we know of certain individuals who are incapable of developing a sense of justice, such as certain autistic persons. We certainly owe them something. But again to try to give them what justice would advise – and expect

---

619 *TJ*, p. 509/446. This sentence occurs within a paragraph discussing the rights of children. But it appears to be perfectly appropriate here. The reference to "fortuitous circumstances" is to being a child in the well-ordered society when the members of that society reflect on the thought-experiment of the original position.

620 Dudley Knowles has insisted to me that he believes that Rawls was most likely being moved by moral compunctions to extend the scope of justice at this point, and adds "Good for him." I say good for him as well, but this does not mean that such an extension would fit very well with the rest of his theory. From the most coherent interpretation of Rawls's theory (not his overall thought and attitudes), I believe it does not fit at all, and that Rawls became increasingly aware of this over the course of his career.

this from them – would carry a risk of putting them and others in danger.

The second paragraph, together with the latter sentence I have quoted, suggest that an inability to develop a sense of justice can be viewed in certain cases as a contingent deprivation. This is the third possible case which was left aside in subsection 15.5 D. What do we now say about those who *had* a capacity for the sense of justice, but whose upbringing has meant that this disposition now cannot be realised? This occurrence appears to be a morally arbitrary matter. But on the other hand, this person is now outside the scope of relationships of reciprocity. An obvious tension arises in respect to the considerations for restricting justice to those with the capacity for a sense of justice which we have just proposed.

In response, I note that such deprivations often represent non-ideal circumstances. I shall not explore what non-ideal theory would require for such cases as this. In the ideal case, any such loss of one's capacity to develop the moral powers would most likely not occur. If it did, this would be due to misfortune in otherwise favourable circumstances and just institutions. This limits, but does not eliminate our problems, but I leave aside how they might be resolved for some other time.

I conclude, then, that there are no sufficient conditions to be owed justice, on the most defensible variation of Rawls's theory, apart from possessing the capacity for the moral powers. Practical considerations may lead us to extend the scope of justice further in a legal or conventional sense. But this is not to truly extend it further in a moral sense. Given that there are no further sufficient conditions to be owed justice, having the capacity for the moral powers is both necessary and sufficient to be owed justice.

Section 18: The Demands of Political Justice

In this chapter, I have attempted to develop an account of the scope of justice roughly along the lines taken by Rawls. I have attempted to clarify some of my moral intuitions regarding this kind of approach, in order to present it in a more powerful form. I have also tried to point to elements in Rawls's own theory which suggest that he should be sympathetic to such revisions. Of course, there is no way of knowing how much sympathy he would have shown. I recognise that there is still a lot to be clarified – not least in continuing to work out what the various aspects of Rawls's theory commit him to regarding this matter. But I put these thoughts out, much like those in chapters 2 and 4, to any Rawlsian who would wish to develop them.

But of late, and perhaps even earlier, I have begun to doubt that *I* could ever

endorse this project, and indeed the rest of the Rawlsian political liberal project, myself. Or, more precisely, if I could ever wholeheartedly endorse it, in the way I take most people think it should be endorsed, i.e. as giving us an account of political legitimacy, and the limits of state action.[621] I believe I should put down these reflections here, so that the reader is aware of my dissatisfaction with some of the possible implications which follow from accepting the account of the scope of justice given above, and the ideas of political liberalism.

To lead on to my worry, consider the following passage from *Political Liberalism*.

> We may ask whether justice can be extended to our relations to animals and the order of nature ... In [this] case we start from the status of adult citizens and proceed subject to certain constraints to obtain a reasonable law [for] the claims of animals and the rest of nature; this has been the traditional view of Christian ages. Animals and nature are seen as subject to our use and wont. This has the virtue of clarity and yields some kind of answer. There are numerous political values here to invoke: to further the good of ourselves and future generations by preserving the natural order and its life-sustaining properties; to foster species of animals and plants for the sake of biological and medical knowledge with its potential applications to human health; to protect the beauties of nature for purposes of public recreation and the pleasures of a deeper understanding of the world. The appeal to values of this kind gives what many have found a reasonable answer to the status of animals and the rest of nature.
>
> Of course, some will not accept these values as alone sufficient to settle the case. Thus, suppose our attitude towards the world is one of natural religion: we think it utterly wrong to appeal solely to those values, and others like them, to determine out relations with the natural world. To do that is to see the natural order from a narrowly anthropocentric point of view whereas human beings should assume a certain stewardship towards nature and give weight to an altogether different family of values. In this case our attitude might be much the same as those who reject abortion on theological grounds. Yet there is this

621 That this is the ambition of political liberalism is clear from many elements – see especially *PL* on liberal legitimacy (pp. 135—137), and the priority of right (pp. 173—176). Quong's (2011) introduction to his book-length defence of political liberalism makes is clear that the primary interest of political liberalism is to set the ultimate bounds on the actions of the state. See esp. pp. 1—2

> important difference: the status of the natural world and our proper relation to it is not a constitutional essential or a basic question of justice as these questions have been specified. It is a matter in regard to which citizens can vote their nonpolitical values and try to convince other citizens accordingly. The limits of public reason do not apply.[622]

We should make some initial observations. The first is that the status Rawls attributes to animals here is less than in the earlier philosophy. As I mentioned in subsection 16.2 Rawls there held that we possess duties of compassion and humanity towards animals. Rawls must presumably view these duties as components of a comprehensive doctrine, as they do not seem to be referred to here.

The second observation is that, given Rawls's assumptions, I cannot see why he should not say roughly similar things about those without the capacity for one or both of the moral powers. Remember that the majority of such people are not dangerous psychopaths.

Rawls's later position, then – and perhaps also the earlier one, if the commitments of public justification are thought through – leaves those individuals without the moral powers lacking any intrinsic political standing. How are we then allowed to act, if we share a society with a culture which would deny animals, or those persons without the moral powers, *any* moral standing? We cannot legislate against them doing what they like. It would be illegitimate. The members of such a culture might perfectly abide by public reason. We cannot intervene to stop their behaviour. This would be illegal. All we can do is try to convince them, non-publicly, to change their ways.

I see no obvious way in which Rawls, or indeed any sufficiently similar contractualist or public justification liberalism, can go against this conclusion, or alter their theory so as to avoid it. Hence, for example, returning to a topic in subsection 15.5 A above, I find it difficult to see how Rawls can fit human rights – which I take to be rights which apply simply to all members of *Homo Sapiens* – into Justice as Fairness in anything more than an ad hoc manner. For the individuals who cannot stand in a reciprocal moral relationship with moral persons simply cannot have their own political standing: how can they be part of the original contract?[623]

---

622 *PL*, p. 245—246

623 Other contractarian proposals for dealing with those who are unable to express the Rawlsian moral powers, or who cannot contribute to the upkeep of the basic structure of society in the way Rawls assumes (subsection 15.2), but who can nevertheless reciprocate in some fashion, may face a lessened version of this problem. Nevertheless, they may still threatened by it (as examples, see Silvers and Francis (2005) and Hartley (2009)). I do not investigate these further matters here.

I do not think this is necessarily a problem for the theory as an analysis of justice. I view it as a problem if the theory is taken to set ultimate limits on legitimate political action. For these are simply things we must be willing to legislate regarding. I have reflected as long and hard about this as about anything. I believe that we cannot simply treat animals as we please, and that we should not allow others to do so either. The same goes for persons who are not capable of moral personality. My point is that we cannot simply put these considerations aside, and see them as a non-public matter, if we happen to be in a well-ordered society in which others in the society deny the interests of these individuals (animals and humans). The very fact that this can be seen as even *potentially* a non-public, non-political matter is what is problematic.

To make vivid what is at issue here, consider the following scenario. If we lived in a world in which everyone was born with a moral capacity which could be realised, and there were no animals, things would be different. In John W. Campbell Jnr's[624] 1951 novella *The Moon is Hell*, a spaceship crash-lands on the dark side of the Moon. But, the astronauts work out how to synthesize all of the oxygen and food they need from the inorganic matter of the Moon. A society descended from those astronauts could organise itself perfectly according to the requirements of a liberal political conception without worrying about animals, at least. But we are faced with the circumstances we have. No appeal to the inevitability of reasonable pluralism, and the limitations it brings on public justification, can eliminate the fact that there are individuals who fall outside the scope of justice, and hence political standing, as these are understood by Rawls.

I believe this represents a limit on the extent to which anyone should subscribe to political liberalism, and to Rawlsian contractualism. I also believe it represents a limit to the extent that any *government* can so subscribe — and this is problematic, as political liberalism incorporates what is meant to be itself a full account of liberal legitimacy. At this point, I feel I simply have to say that it must be an incomplete account. I simply do not believe that governments can ignore the interests of the individuals I have identified. And no amount of considering the importance of political reasonableness, or the burdens of judgement, is going to make me think otherwise. It is not that I do not grasp that there could be considerations to be balanced here. If certain ritual practices mean that domestic animals must be slaughtered in certain ways, I can recognise that there are things to be said on both sides (note that in the considerations presented here, I have not taken myself to be arguing for vegetarianism or veganism necessarily – much as I think vegetarianism pretty

624 John W. Campbell Jnr. was an American science fiction author, and editor of the historically important *Astounding Science Fiction*.

much holds all the cards in its moral debate). But the strictures of public reasoning appear to prevent me from saying that there is anything on the side of the animals which does not derive from my own interest in them. And this is the wrong way to view the matter, so I unavoidably believe.

I realise that these comments come to less than full philosophical argument. But they do relate a commitment which, no matter how hard I tried, I do not think I could ever give up. As *it is not mine to give up* – these animals and persons really do exist here in the world with us. We can attempt to develop a theory to its most charitable form. But when we are as sure as we can be that we are not being self-serving or blinkered, we cannot ignore what we conscientiously take to have value of its own.

Finally, I might comment on the question as to whether this problem could be remedied by revising the scope of justice. I am unsure, as I am unsure what the content of justice is. However, what I firmly believe is that we cannot simply revise our conception of justice to incorporate whatever we might like. It may be that, in the final analysis, a government which passes anti-cruelty legislation and healthcare legislation to protect individuals with the relevant severe mental or neurological conditions, given a certain public culture and certain pattern of comprehensive doctrines, is behaving unjustly. But if this is true, then it just goes to show how it is sometimes right for a government to behave unjustly, and how the scope of justice does not always link to the scope of justified government action.[625]

625 In this conclusion, I have obviously been influenced by work by Cohen (2008) esp. chapters 6 and 7. I must also acknowledge my debt to Steiner (1994) through my long reflections on what I did not agree with about his theory, and how I could respond. His theory I take to have the similar problems as Rawls's, only they are even more severe. The same sentiments which move me are presumably those that move Nussbaum (2006), with one difference – I am less convinced than she that the importance of animals, and some of the severally impaired, is a matter of justice. But then, I believe that thinking that unless we can show an issue to be a matter of justice, then we have nothing strong enough to say against our opponents is a regrettable kowtowing to a certain prominent trend in political and legal thought.

Epilogue

I have laid out the roles of moral psychology within Rawls's theory. Moral psychology is to demonstrate how Rawls's principles, and the society which embodies them, is realisable and stable. Through playing this role, moral psychology is then able to play a further two roles. It is able to play its part in justifying the Rawlsian society, through showing it is not a futile ambition to attempt to realise that society, and also through showing that such a society is comparatively more stable than societies based on different principles. If we understand Rawls's moral theory to be, taken as a whole, an account of the morality of the just society, then given that the moral psychology of the persons in that society is part of what can be properly called the morality of that society, then moral psychology is constitutive of (at least part of) morality (from within the perspective of Rawls's theory). It is the possession of (the capacity for) Rawls's moral psychology which identifies those who fall within the scope of justice. These roles, and the overall place in Rawls's theory which moral psychology occupies, does not significantly change between Rawls's earlier and later work.

I should like to end this work with some comments on the significance of my work, both in itself and within the broader context of moral philosophy. The ideas that follow may have been suggested to the reader by the preceding chapters, or they may have not. They are regrettably sketchy – I cannot render them as precise as I would like at this stage, though I believe it would be quite possible to do so.

What overall shape does moral psychology have within Rawls's theory? Given his fundamental normative assumptions, and the shape of his theory, moral psychology then interacts with our moral intuitions, to finalise the contract between our representatives in the original position. The principles agreed to are then for us, as they are ultimately agreed to by us in the right way, and they are between us, as beings without the right moral capacities are excluded. Moral psychology – an element of the full description of human nature – hence completes an account of morality which aims to arise entirely from ourselves, to apply to ourselves, and to be realisable by ourselves.

There are three ways such an account of morality might be rejected. We might reject the account of the capacities of human nature, and reject the idea that human beings can possess the moral powers. Or we might reject the idea that these capacities are most

deeply seated in human nature as it would be realised in a free and equal society, and instead claim, for example, that altruism is. This is to dispute Rawls's claims about human nature at their broadest scope. To dispute Rawls claims at narrower levels would be to accept the broad account of the moral powers, and the basing of our fundamental moral powers on the psychological propensity to reciprocity, but to derive different principles from these presuppositions, on the basis of different fine grained claims about human psychology (plus sociology etc.).

A third and final way to dispute Rawls's account of morality would be to reject his assumption that the content of morality must ultimately be constrained by human nature. Moral psychology would then describe the psychology of moral beings, but without a guarantee at all that the characteristics of human beings match up to these moral beings.

The exegesis and analysis I have presented here can help us to orientate ourselves in investigating the first and second ways in which we might dispute Rawls's account. But it cannot help us decide the third matter, as it has been assumed by the entire analysis. The full, honest, philosophical assessment of the content of morality – how we should act if we are to be moral beings – and whether that content can contain requirements which we cannot act on, is, to my eyes the fundamental issue moral philosophers should address, if they are to understand themselves to be, as I think they should do, those who are interested in morality for its own sake, and for the sake of nothing else. Any posture which presumes the answer to these questions addresses not morality directly, but morality through the lens of partial interest and rhetoric, and runs at least the risk of distorting and admixing our view of morality with other concerns: other concerns including those worthwhile themselves, but the worthless.

Appendix I: Constructivism

Constructivisms are anti-realist but (most usually) objectivist accounts of morality.[626] Rawls's constructivism about justice is a particular variety of constructivism. To begin to outline Rawls's constructivism, we should distinguish between practical and theoretical reason. Rawls writes

> Following Kant's way of making the distinction, we say: practical reason is concerned with the production of objects according to a conception of those objects – for example, the conception of a just constitutional regime taken as the aim of political endeavour – while theoretical reason is concerned with the knowledge of given objects.[627]

In constructivism, then, practical reason *produces* moral principles and moral reasons. In some constructivisms, practical reason also produces prudential principles and reasons, evaluative reasons (and principles), and so on for other normative and evaluative types.[628] In Rawls's constructivism,[629] however, only reasonable principles – which include principles of justice[630] – are produced.[631] In Rawls's constructivism, then, *principles of practical reason* are rational and reasonable principles. Principles of practical reason, simply put, are the principles that are applied by us in "reasoning about what to do."[632]

Rawls often contrasts constructivism to what he calls rational intuitionism. The debate between constructivism and rational intuitionism,[633] is whether moral principles are

626 Anti-realist positions see morality as mind-dependent, as opposed to mind independent. See the characterisation of constructivism in Shafer-Landau and Cuneo (2007) pp. 79—83. Objectivist views hold that moral requirements are the same for all agents – they do not differ between different individuals or groups of agents. Not all views classified as constructivist are objectivist (i.e. Harman (1975)). Even if one accepts this classification (as not all do, e.g. O'Neill (2003) p. 348), it is true to say that *most* developed constructivisms are objectivist.

627 *PL*, p. 93

628 For example, Street (2008) pp. 208—209 esp fn4, 223—242

629 Rawls's most thorough treatments of his constructivism can be found in *Political Liberalism* and "Kantian Constructivism in Moral Theory" and his account of Kant's constructivism in *Lectures on the History of Moral Philosophy*. I note here that these are the central sources of my understanding of constructivism. I am not a Kant scholar, and do not have extensive knowledge of the wider debate.

630 *PL*, p. 83

631 *TJ*, p. 446/392 is evidence for that principles of rationality are not constructed. On Rawls's reading of Kant this is true as well (*LHMP*, pp. 237, 239).

632 *LP*, p. 87

633 See *PL*, pp. 90—92, *CP*, pp. 343—345 for the basics of Rawls's understanding of rational intuitionism. Rawls's understanding of "rational intuitionism" is overly Platonic. Many contemporary

discovered by theoretical reason, or produced by practical reason.[634] In rational intuitionism practical reason still plays a role. Objects are still created in accordance with our conception of them – namely our actions, and the products of our actions. But in constructivism, not only the actions but the principles which guide those actions are produced.

Practical reason is composed of *conceptions of practical reason* as well as principles of practical reason. Conceptions of practical reason "characterise the agents who [practically] reason and ... specify the context for the problems and questions to which principles of practical reason apply."[635] In Rawls's constructivism, such conceptions include the conception of persons as reasonable and rational, free and equal, and the conception of the well-ordered society governed by a public conception of justice. Conceptions of practical reason are not constructed.[636] But they are said to "arise" and to be "appropriate" because they complement the principles of practical reason. The "principles do not apply themselves, but are used by us in forming our intentions ... and plans ..., in our relations with other persons" such that " without conceptions of society and person, the principles of practical reason would have no point, use, or application."[637] This last point is obscure. How I interpret it is that conceptions of practical reason are not constructed. But they only acquire normative authority if we are able to construct principles which compliment them. If we cannot, then they are empty conceptions. They fail to describe our actual faculties of practical reasoning, and hence do not refer to

---

intuitionists would also call themselves naturalists, in contrast with Rawls's usage over these passages. The central intuitionist claims are simply that we know moral principles through *a priori* theoretical (not practical) reasoning, and also that moral concepts cannot be reduced down to those of the natural sciences. I am indebted to discussions with Robert Cowan on these matters.

634 *PL*, pp. 91—93, 96

635 *PL*, p. 107

636 *PL*, p. 108. Note that when Rawls also says here that the "principles of practical reason" are not constructed, he surely can't mean the principles of justice and the other reasonable principles. I see two ways out. He may have simply meant the rational principles. Alternatively, in *LP*, p. 86—87 incl. fn33, Rawls comments that "at no point are we deducing the principles of right and justice, or decency, or the principles of rationality, from a conception of practical reason in the background." He confesses that "there are many places in [*PL*] where I gave the impression that the content of the reasonable and the rational is derived from the principles of practical reason." Instead, he now only claims that Justice as Fairness gives "content to an idea of practical reason" and that the specification of the normative ideas of the reasonableness and rationality "are not deduced, but enumerated and characterised in each case." I do not think anything of significance I say hinges on this change. Though to say for sure would take more examination, his most obvious concern here seems to be to distinguish his view from Kant's. I conjecture this may all be in order to ensure that the requirements of political liberalism are met. Hence "an idea" of practical reason, rather than just practical reason straight up. I am at a loss as to how to make sense of the distinction made here between principles of practical reason, and the content of the reasonable and rational. As far as the text of *PL* reads, the principles of practical reason aren't meant to be prior to this content, but are rather the same thing. Perhaps by content he is thinking of reasons of reasonableness and rationality, which he *appears to* distinguish from principles (*PL*, pp. 121—122). Street (2008) pp. 210—211 manages to produce a very coherent reading from this observation.

637 *PL*, pp. 107—108

anything.

Constructivist theories usually hold that moral or political principles are to be conceived as the outcome of a "procedure of construction" or a construction procedure.[638] The original position is an example of such a procedure, as are contractarian procedures in general. Construction procedures can also be non-contractarian, such as Ideal Observer Theory.[639] Such procedures of construction are "*based essentially* on practical reason and not on theoretical reason."[640] What this means is that they are *assembled*[641] out of conceptions of practical reason – which as I have said are not constructed or produced themselves – and then produce principles of practical reason.

Theoretical reason also has a role to play in assembling the original position. Theoretical reason "shapes the belief and knowledge of the rational persons who have a part in the construction; and these persons also use their general capacities of reasoning, inference and judgement in selecting principles of justice."[642] In addition, the parties are provided with factual data and knowledge for theoretical reason to work upon.[643]

Rawls's assumption that conceptions of practical reason are not constructed mirrors his reading of Kant.[644] Onora O'Neill questions this reading, holding that Kant can be read as holding the conceptions of practical reasoning to also be constructed by practical reasoning.[645] It is beyond my expertise to adjudicate. But it is clear that Rawls follows his reading of Kant in developing his own version of constructivism.

Rawls proposes two constructivisms in his work. The first is his Kantian Constructivism. This is a moral constructivism which is meant to provide the metaethical basis of a comprehensive moral conception of justice. The second is his Political Constructivism. This provides the metaethical basis only for a political conception of justice. Structurally, the two constructivisms are the same, except that due to the requirements of political liberalism, political constructivism only claims that its political conception is constructed from its own standpoint, not from the standpoint of comprehensive moral doctrines.[646] Kantian constructivism, by contrast, claims that justice

638 *PL*, pp. 89—90, 93. For further characterisations compatible with Rawls's, see Cohen (2008) pp. 274—276, Barry (1989) pp. 264—82, 348—53. Note I say most: for example, Street (2008) is not committed to this.
639 Cohen (2008) p. 275 observes this. This appears to make good sense of *TJ*, pp. 183—189/160—165
640 *PL*, p. 93. My emphasis
641 *PL*, p. 108, also 103
642 *PL*, p. 108
643 *PL*, pp. 121—123
644 *LHMP*, pp. 239—240, 253—261, 268—271
645 See O'Neill (2003) pp. 356—361. See further O'Neill (1989) chapter 1
646 See, for example, *PL*, pp. 119—120, 128—129

is constructed from a comprehensive perspective.[647] On the distinction between political and comprehensive conceptions, see subsection 12.1

Appendix II: Psychological Tendencies to Reciprocity and Altruism

Rawls holds that human moral psychology is governed by, amongst other tendencies, psychological tendencies towards altruism, and psychological tendencies towards reciprocity. Psychological principles of reciprocity state that human beings have a tendency to respond in kind to how others have acted towards them. Psychological principles of altruism state that human beings have a tendency to care for the good of others, independent of their own good.

Rawls develops an account of moral development in which moral development occurs in stages, following the Cognitive-Developmental School of Jean Piaget and Lawrence Kohlberg.[648] Rawls postulates three such stages: the initial stage of moral sensibility developed by young children,[649] a mediate stage of moral development which occurs between later childhood and early adulthood, as we enter various associations in civil society,[650] and a final stage at which we acquire an attachment to acting morally apart from our ties to particular others.[651] At each stage, moral development occurs due to our good being cared for by those around us, because the reasons for the need to act morally are explained clearly, and because the forms of life displayed by those around us are admirable and display human virtue and excellence.[652] Transition between these stages is roughly governed by psychological principles of reciprocity.[653] These represent a more general psychological tendency towards reciprocity – "a tendency to answer in kind."[654]

Rawls does not deny that human beings also have psychological tendencies to altruism.[655] But he speculates that these are less powerful or pervasive, and hence less appropriate as a foundation for a moral psychology.[656]

---

647 See, for example, *CP*, pp. 353—356

648 See *TJ*, p. 461/404 fn8. For criticisms of the Cognitive-Developmental School, see Flanagan (1991) chapters 7 and 8. Rawls's claims and requirements for his moral psychology are less than those psychologists, so I am not convinced that problems for the Cognitive-Developmental School necessarily spell trouble for him.

649 *TJ*, pp. 462—467/405—409

650 *TJ*, pp. 467—472/409—413

651 *TJ*, pp. 472—479/414—419

652 *TJ*, pp. 498—499/436

653 *TJ*, pp. 490—491/429—430

654 *TJ*, p. 494/433. See also *JF*, pp. 195—196

655 E.g. *TJ*, p. 486

656 E.g. *TJ*, p. 500—501/437—438

General psychological tendencies such as reciprocity and altruism do not represent moral psychologies in themselves. Moral psychologies, as stressed in subsection 5.2, presuppose moral principles. Without this, psychological tendencies such as reciprocity and altruism can be orientated towards evil as easily as good.[657]

657 See, for example, *TJ*, pp. 190/166, for altruism. Similar scenarios of misplaced reciprocity can be reconstructed from pp. 472—473/413—414.

Bibliography

Anderson, Elizabeth (2010) "Justifying the Capabilities Approach to Justice" Chapter 4 in Brighouse, Harry, and Robeyns, Ingrid (eds.) *Measuring Justice: Primary Goods and Capacities*, Cambridge University Press, Cambridge

Baldwin, Thomas (2008) "Rawls and Moral Psychology" pp. 247—270 in Shafer-Landau, Russ (ed.) *Oxford Studies in Metaethics: Volume 3*, Oxford University Press, Oxford

Barry, Brian (1989) *A Treatise on Social Justice Volume I: Theories of Justice*, Harvester-Wheatsheaf, London

Barry, Brian (1995a) *A Treatise on Social Justice Volume II: Justice as Impartiality*, Clarendon Press, Oxford

Barry, Brian (1995b) "John Rawls and the Search for Stability" *Ethics*, 105, pp. 874—915

Bates, Stanley (1974) "The Motivation to be Just" *Ethics*, 85, pp. 1—17, reprinted in Weithman, Paul J. (ed.) *The Philosophy of Rawls: A Collection of Essays – Volume 4: Moral Psychology and Community*, pp. 67—83, Garland Publishing Inc., New York

Bonotti, Matteo (unpublished) "Political Liberalism, Distributive Justice and the Relationship between State and Religion" presented at Political Philosophy Seminar, University of Stirling (2011)

Buchanan, Allen (1990) "Justice as Reciprocity Versus Subject-Centred Justice" *Philosophy and Public Affairs*, 19, pp. 227—252

Carter, Ian (2011) "Respect and the Basis of Equality" *Ethics*, 121, pp. 538—571

Ci, Jiwei (2006) *The Two Faces of Justice*, Harvard University Press, Cambridge Mass.

Cohen, G.A. (1989) “On the Currency of Egalitarian Justice” *Ethics*, 99, pp. 906—944

Cohen, G.A. (2008) *Rescuing Justice and Equality*, Harvard University Press, Cambridge Mass.

Cohen, Joshua (2002) “Taking People as They Are?” *Philosophy & Public Affairs*, 30, No. 4, pp. 363—386

Cushman, Fiery, Young, Liane, and Greene, Joshua D. (2010) “Multi—System Moral Psychology”, Chpt. 2 in Doris, John M. and the Moral Psychology Research Group (eds.), *The Moral Psychology Handbook*, Oxford University Press, Oxford

Darwall, Stephen L. (1977) “Two Kinds of Respect” *Ethics*, 88, pp. 36—49

Deigh, John (1983) “Shame and Self-Esteem: a Critique” *Ethics*, 93, pp. 225—245

Deigh, John (1996) *The Sources of Moral Agency: Essays in Moral Psychology and Freudian Theory*, Cambridge University Press, New York

DePaul, Michael R. (1986) “Reflective Equilibrium and Foundationalism” *American Philosophical Quarterly*, Vol. 23, No. 1, pp. 59—69

Dillon, Robin (2010) “Respect” *The Stanford Encyclopedia of Philosophy*, Zalta, Edward, N. (ed.), URL = <http.//plato.stanford.edu/entries/respect>

Doppelt, Gerald (2009) “The Place of Self-Respect in a Theory of Justice” *Inquiry*, Vol. 52, No. 2, pp. 127—154

Dworkin, Ronald (1981) “What is Equality? Part 2: Equality of Resources” *Philosophy and Public Affairs*, 10, pp. 283—345

Ebertz, Roger, P. (1993) “Is Reflective Equilibrium a Coherentist Model?” *Canadian Journal of Philosophy*, Vol. 23, No. 2, pp. 193—214

Estlund, David (1996) “The Survival of Egalitarian Justice in John Rawls's *Political Liberalism*” *Journal of Political Philosophy*, 4(1), pp. 68—78, reprinted in Kukathas, Chandran (ed.) *John Rawls: Critical assessments of leading political philosophers; Volume IV: **Political Liberalism** and **The Law of Peoples***, pp. 380—391, Routledge, London.

Estlund, David (1998) “The Insularity of the Reasonable: why political liberalism must admit the truth” *Ethics*, 108(2), pp. 252—275, reprinted in Kukathas, Chandran (ed.) *John Rawls: Critical assessments of leading political philosophers; Volume IV: **Political Liberalism** and **The Law of Peoples***, pp. 86—109, Routledge, London.

Eyal, Nir (2005) “'Perhaps the most important primary good': self-respect and Rawls's principles of justice” *Politics, Philosophy & Economics*, 4(2), pp. 195—219

Flanagan, Owen (1991) *Varieties of Moral Personality: Ethics and Psychological Realism*, Harvard University Press, Cambridge, Mass.

Flanagan, Owen (1996) “Ethics Naturalized: Ethics as Human Ecology” in May, Larry, Friedman, Marilyn, and Clark, Andy (eds.) (1996) *Mind and Morals: Essays on Ethics and Cognitive Science*, MIT Press, Cambridge, Mass.

Flanagan, Owen, Sarkissian, Hagop, and Wong, David (2008) “Naturalizing Ethics” in Sinnott-Armstrong, Walter (ed.) *Moral Psychology: Volume 1*, MIT Press, Cambridge Mass.

Frankfurt, Harry (1971) “Freedom of the will and the concept of a person” *Journal of Philosophy* 68, pp. 5—20

Frazer, Michael L. (2007) “John Rawls: Between Two Enlightenments” *Political Theory*, Vol. 35, No. 6, pp. 756—780

Freeman, Samuel (2003) “Congruence and the Good of Justice” in Freeman, Samuel (ed.) *The Cambridge Companion to Rawls*, Cambridge University Press, New York

Freeman, Samuel (2006) “Book Review: Frontiers of Justice: The Capabilities Approach

vs. Contractarianism" *Texas Law Review*, 85, pp. 385—430

Freeman, Samuel (2007a) *Rawls*, Routledge, London

Freeman, Samuel (2007b) *Justice and the Social Contract: Essays on Rawlsian Political Philosophy*, Oxford University Press, New York

Gaus, Gerald F. (1996) *Justificatory Liberalism: An Essay on Epistemology and Ethical Theory*, Oxford University Press, New York

Gaus, Gerald F. (2003) *Contemporary Theories of Liberalism: Public Reason as a Post—Enlightenment Project*, Sage Publications, London

Gaus, Gerald (2011) *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*, Cambridge University Press, New York

Gibbard, Allan (1982) "Human Evolution and the Sense of Justice" pp. 31—46 in Uehling, Jnr., Theodore E. and Wettstein, Howard K. (eds.), *Social and Political Philosophy, Midwest Studies in Philosophy 7*, reprinted in Weithman, Paul J. (ed.) *The Philosophy of Rawls: A Collection of Essays – Volume 4: Moral Psychology and Community*, pp. 85—100 Garland Publishing Inc., New York

Gibbard, Allan (1991) "Constructing Justice" *Philosophy and Public Affairs*, 20, pp. 264—279

Hare, R.M. (1975) "Rawls' Theory of Justice" in Daniels, Norman (ed.) *Reading Rawls: Critical Studies of **A Theory of Justice***, Basil Blackwell, Oxford

Hart, H.L.A. (1973) "Rawls on Liberty and its Priority" *University of Chicago Law Review*, 40, pp. 534—555, reprinted in Daniels, Norman (ed.) *Reading Rawls: Critical Studies of **A Theory of Justice***, Basil Blackwell, Oxford

Hartley, Christie (2009) "An Inclusive Contractualism: Obligations to the Mentally Disabled" chapter 5 in Brownlee, Kimberley and Cureton, Adam (eds.) *Disability and Disadvantage*, Oxford University Press, Oxford.

Hauser, Marc D., Young, Liane, and Cushman, Fiery (2008) “Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions” pp. 107—143 in Sinnott-Armstrong, Walter (ed.) *Moral Psychology: Volume 2*, MIT Press, Cambridge Mass.

Haybron, Dan (2008) *The Pursuit of Unhappiness*, Oxford University Press, Oxford

Held, Virginia (2006) *The Ethics of Care: Personal, Political, and Global*, Oxford University Press, New York

Hobbes, Thomas (1651) *Leviathan, or The Matter, Forme, & Power of a Common-Wealth Ecclesiasticall and Civill*, Tuck, Richard (ed.) (1996) Revised Student edition, Cambridge University Press, Cambridge

Hope, Simon (2010) “The Circumstances of Justice” *Hume Studies*, Vol. XXXVI, No. 2, pp. 125—148

Hubin, D. Clayton (1979) “The Scope of Justice” *Philosophy & Public Affairs*, 9, No. 1, pp. 3—24

Hume, David (1739) *A Treatise of Human Nature: being an attempt to introduced the experimental method of reasoning into moral subjects*, edited by Norton, David Fate, and Norton, Mary J. (2000) Oxford Philosophical Texts edition, Oxford, New York

Hume, David (1751) *Enquiry concerning the Principles of Morals* in *David Hume: Enquiries concerning Human Understanding and Concerning the Principles of Morals*. Selby-Bigge, L.A. And Nidditch, P.H. (1975) Clarendon Press, Oxford.

Kearns, Deborah (1983) “A Theory of Justice – and Love; Rawls on the Family” *Politics* (Australasian Political Studies Association Journal) 18 (2), pp. 36—42

Kelly, Erin (2010) “Equal Opportunity, Unequal Capacity” chpt 3 in Brighouse, Harry, and Robeyns, Ingrid (eds.) *Measuring Justice: Primary Goods and Capacities*, Cambridge University Press, Cambridge

Kittay, Eva Feder (1999) *Love's Labour: Essays on Women, Equality, and Dependency*, Routledge, Abingdon

Klosko, George (1994) "Rawls's Argument from Political Stability" *Columbia Law Review*, Vol. 94, pp. 1882—1897.

Krasnoff, Larry (1998) "Consensus, Stability and Normativity in Rawls's Political Liberalism" *The Journal of Philosophy*, Vol. XCV, No. 6, pp. 269—292

Krause, Sharon R. (2008) *Civil Passions: Moral Sentiment and Democratic Deliberation*, Princeton University Press, Princeton

Labukt, Ivar (2009) "Rawls on the Practicality of Utilitarianism", *Politics, Philosophy and Economics*, 8(2) pp. 201—221

Larmore, Charles (1990) "Political Liberalism" *Political Liberalism*, Vol. 18, No. 3, pp. 339—360

May, Larry, Friedman, Marilyn, and Clark, Andy (eds.) (1996) *Mind and Morals: Essays on Ethics and Cognitive Science*, MIT Press, Cambridge, Mass.

Mendus, Susan (2002) *Impartiality in Moral and Political Philosophy*, Oxford University Press, Oxford

Mill, John Stuart (1863) "Utilitarianism" reprinted in Gray, John (ed.) *On Liberty and Other Essays*, (1998) Oxford World's Classics, Oxford University Press, Oxford

McClennen, Edward F. (1989) "Justice and the Problem of Stability" *Philosophy and Public Affairs* 18, pp. 3—30 reprinted in Weithman, Paul J. (ed.) *The Philosophy of Rawls: A Collection of Essays – Volume 4: Moral Psychology and Community*, pp. 85—100 Garland Publishing Inc., New York

Mikhail, John (2011) *Elements of Moral Cognition: Rawls's Linguistic Analogy and the Cognitive Science of Moral and Legal Judgement*, Cambridge University Press, New York

Murphy, Liam B. (1999) “Institutions and the Demands of Justice” *Philosophy and Public Affairs*, 27, No. 4, pp. 251—291

Nussbaum, Martha C. (1990) *Love's Knowledge: Essays on Philosophy and Literature*, Oxford University Press, New York

Nussbaum, Martha C. (2003) “Rawls and Feminism” in Freeman, Samuel (ed.) *The Cambridge Companion to Rawls*, Cambridge University Press, New York

Nussbaum, Martha C. (2006) *Frontiers of Justice: Disability, Nationality, Species Membership*, The Belknap Press of Harvard University Press, Cambridge Mass.

Okin, Susan Moller (1989) “Reason and Feeling in Thinking about Justice” *Ethics*, 99, pp. 229—249

Okin, Susan Moller (1994) “*Political Liberalism*, Justice, and Gender” *Ethics*, 105, pp. 23—43

Parfit, Derek (1984) *Reasons and Persons*, Oxford University Press, Oxford

Perry, John (2008) *Personal Identity: second edition*, University of California Press, Berkeley and Los Angeles

Prinz, Jesse (2007) *The Emotional Construction of Morals*, Oxford University Press, Oxford

Pritchard, Michael S. (1977) “Rawls's Moral Psychology” *Southwestern Journal of Philosophy*, 8, pp. 59—72

Pogge, Thomas (2007) *John Rawls: His Life and Theory of Justice*, Oxford University Press, Oxford

Pogge, Thomas (2010) “A Critique of the Capability Approach” chapter 1 in Brighouse, Harry, and Robeyns, Ingrid (eds.) *Measuring Justice: Primary Goods and Capacities*,

Cambridge University Press, Cambridge

Quong, Jonathan (2007) “Contractualism, Reciprocity and Egalitarian Justice” *Politics, Philosophy and Economics*, 6(1), pp. 75—105

Quong, Jonathan (2011) *Liberalism Without Perfection*, Oxford University Press, Oxford

Rawls, John (1971) *A Theory of Justice* (original edition), Belknap Press of Harvard of University Press, Cambridge, Mass.

Rawls, John (1999) *Collected Papers*, Freeman, Samuel (ed.), Harvard University Press, Cambridge, Mass.

Rawls, John (1999) *The Law of Peoples*, Harvard University Press, Cambridge, Mass.

Rawls, John (2000) *Lectures in the History of Moral Philosophy*, Herman, Barbara (ed.), Harvard University Press, Cambridge, Mass.

Rawls, John (2001) *Justice as Fairness: A Restatement*, Kelly, Erin (ed.), The Belknap Press of Harvard University Press, Cambridge, Mass.

Rawls, John (2005) *Political Liberalism* (expanded edition), Columbia University Press, New York

Rawls, John (2007) *Lectures on the History of Political Philosophy*, Freeman, Samuel (ed.), The Belknap Press of Harvard University Press, Cambridge, Mass.

Raz, Joseph (2003a) “The Claims of Reflective Equilibrium” in Kukathas, Chandran (ed.) *John Rawls: Critical assessments of leading political philosophers; Volume I: Foundations and Method*, Routledge, London, reprinted from *Inquiry*, (1982), 25, pp. 307—330,

Raz, Joseph (2003b) “Facing Diversity: the case of epistemic abstinence” Kukathas, Chandran (ed.) *John Rawls: Critical assessments of leading political philosophers; Volume IV: **Political Liberalism** and **The Law of Peoples***, Routledge, London, reprinted from *Philosophy and Public Affairs*, (1990), 19(1), pp. 3—46,

Richardson, Henry S. (1994) *Practical Reasoning about Final Ends*, Cambridge University Press, New York

Richardson, Henry S. (2006) "Rawlsian Social-Contract Theory and the Several Disabled" *Journal of Ethics*, 10, pp. 419—462

Roedder, Erica, and Harman, Gilbert (2010) "Linguistics and Moral Theory" in Doris, John M. and the Moral Psychology Research Group, *The Moral Psychology Handbook*, Oxford University Press, Oxford

Sachs, David (1981) "How to distinguish Self-Respect from Self-Esteem" *Philosophy and Public Affairs*, 10, pp. 346—360, reprinted in Weithman, Paul J. (ed.) *The Philosophy of Rawls: A Collection of Essays – Volume 4: Moral Psychology and Community*, pp. 85—100 Garland Publishing Inc., New York

Sandel, Michael J. (1998) *Liberalism and the Limits of Justice: Second Edition*, Cambridge University Press, New York

Scanlon, T.M. (1992) "The Aims and Authority of Moral Theory" *Oxford Journal of Legal Studies*, Vol. 12, No. 1

Scanlon, T.M. (1998) *What We Owe To Each Other*, The Belknap Press of Harvard University Press, Cambridge Mass.

Scanlon, T.M. (2003) "Rawls on Justification", in Freeman, Samuel (ed.) *The Cambridge Companion to Rawls*, Cambridge University Press, New York

Schwartz, Adina (1973) "Moral Neutrality and Primary Goods" *Ethics*, 83, pp. 294—307

Schwarzenbach, Sibyl A. (2009) *On Civic Friendship: Including Women in the State*, Columbia University Press, New York

Scheffler, Samuel (2003) "Rawls and Utilitarianism", in Freeman, Samuel (ed.) *The Cambridge Companion to Rawls*, Cambridge University Press, New York

Sen, Amartya (1980) “Equality of What?” in McMurrin, S. (ed.) *The Tanner Lectures on Human Values*, I: pp. 195—220

Simmons, A. John (2010) “Ideal and Nonideal Theory” *Philosophy and Public Affairs*, 38, No. 1, pp. 5—36

Shafer—Landau, Russ and Cuneo, Terence (2007) *Foundations of Ethics: An Anthology*, Blackwell Publishing, Oxford

Shklar, Judith N. (1989) “The Liberalism of Fear” chapter 1 in Rosenblum, Nancy, L. (ed.) *Liberalism and the Moral Life*, Harvard University Press, Cambridge Mass.

Silvers, Anita, and Francis, Leslie Pickering (2005) “Justice though Trust: Disability and the “Outlier Problem” in Social Contract Theory” *Ethics*, 116, pp. 40—76

Simmons, A. John (2010) “Ideal and Nonideal Theory” *Philosophy and Public Affairs*, 38, No. 1, pp. 5—36

Sinnott—Armstrong, Walter (ed.) (2008) *Moral Psychology; Volume 1: The Evolution of Morality: Adaptations and Innateness; Volume 2: The Cognitive Science of Morality: Intuition and Diversity; Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge, Mass.

Sidgwick, Henry (1907) *The Methods of Ethics*, Macmillan & Co., London

Solomon, Robert C. (1995) *A Passion for Justice: Emotions and the Social Contract*, Rowman and Littlefield Publishers, Inc., Lanham, Maryland

Stark, Cynthia A. (2007) “How to Include the Severally Disabled in a Contractarian Theory of Justice” *The Journal of Political Philosophy*, Vol. 15, No. 2, pp. 125—145

Stark, Cynthia A. (2009) “Contractarianism and Cooperation” *Politics, Philosophy & Economics*, 8(1), pp. 73—99

Steiner, Hillel (1994) *An Essay on Rights*, Blackwell, Oxford

Taylor, Gabriele (1985) *Pride, Shame and Guilt: Emotions of self-assessment*, Clarendon Press, Oxford

Terzi, Lorella (2010) "What Metric of Justice for Disabled People?" chapter 7 in Brighouse, Harry, and Robeyns, Ingrid (eds.) *Measuring Justice: Primary Goods and Capacities*, Cambridge University Press, Cambridge

Thomas, Larry (Lawrence) L. (1977—78) "Rawlsian Self-Respect and the Black Consciousness Movement" *Philosophical Forum*, 9, pp. 303—314, reprinted in Weithman, Paul J. (ed.) *The Philosophy of Rawls: A Collection of Essays – Volume 4: Moral Psychology and Community*, pp. 85—100 Garland Publishing Inc., New York

Thomas, Lawrence (1995) "Self-Respect: Theory and Practice" chapter 13 in Dillon, Robin (ed.) *Dignity, Character, and Self-Respect*, Routledge, New York, reprinted from Harris, Leonard (1983) *Philosophy Born of Struggle: Anthology of Afro-American Philosophy from 1917*, Hunt Publishing Company, Kendall

Titelbaum, Michael G. (2008) "What Would a Rawlsian Ethos of Justice Look Like?" *Philosophy and Public Affairs*, 36, no. 3, pp. 289—322

Tiberius, Valerie and Plakias, Alexandra (2010) "Well-Being" in Doris, John M. and the Moral Psychology Research Group, *The Moral Psychology Handbook*, Oxford University Press, Oxford

Tooley, Michael (1972) "Abortion and Infanticide" *Philosophy and Public Affairs*, 2/1, pp. 37—65

Vanderschraaf, Peter (2006) "The Circumstances of Justice" *Politics, Philosophy & Economics*, 5(3), pp. 321—351

Vanderschraaf, Peter (2011) "Justice as Mutual Advantage and the Vulnerable" *Politics, Philosophy & Economics*, 10(2), pp. 119—147

Wall, Steven (1998) *Liberalism, Perfectionism, and Restraint*, Cambridge University Press, Cambridge

Waldron, Jeremy (1993) "Theoretical Foundations of Liberalism" chapter 2 in *Liberal Rights: Collected Papers 1981—1991*, Cambridge University Press, New York, reprinted from *Philosophical Quarterly*, 37, pp. 127—150

Wenar, Leif (2003) "*Political Liberalism*: an internal critique" Kukathas, Chandran (ed.) *John Rawls: Critical assessments of leading political philosophers; Volume IV:* ***Political Liberalism*** *and* ***The Law of Peoples***, pp. 57—85, Routledge, London, reprinted from *Ethics*, (1995) 106(1), pp. 32—62

Wenar, Leif (2005) "The Unity of Rawls's Work" in Brooks, Thom, and Freyenhagen, Fabian (eds.) *The Legacy of John Rawls*, Continuum, London

Williams, Andrew (1998) "Incentives, Inequality, and Publicity" *Philosophy and Public Affairs*, 27, No. 3, pp. 225—247

Wren, Thomas E. (1991) *Caring about Morality: Philosophical Perspectives in Moral Psychology*, Routledge, London

Wollheim, Richard (1999) *On the Emotions*, Yale University Press, London

Zaino, Jeanne S. (1998) "Self-Respect and Rawlsian Justice" *The Journal of Politics*, Vol. 60, No. 3, pp. 737—753

Zink, James R. (2011) "Reconsidering the Role of Self-Respect in Rawls's *A Theory of Justice*" *The Journal of Politics*, Vol. 73, No. 2, pp. 331—344