

Summer 8-15-2015

Bound to Aim at Good Generally: A Sidgwickian Argument for Rational Impartialism

Tyler Stephen Paytas

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Recommended Citation

Paytas, Tyler Stephen, "Bound to Aim at Good Generally: A Sidgwickian Argument for Rational Impartialism" (2015). *Arts & Sciences Electronic Theses and Dissertations*. 675.

https://openscholarship.wustl.edu/art_sci_etds/675

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Philosophy

Dissertation Examination Committee:

Julia Driver, Chair
Anne Margaret Baxley
Eric Brown
Roger Crisp
Charlie Kurth
Frank Lovett

Bound to Aim at Good Generally: A Sidgwickian Argument for Rational Impartialism

by
Tyler S. Paytas

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2015
St. Louis, Missouri

Table of Contents

Acknowledgements	iii
Abstract	v
Chapter 1: The Partiality Question	1
Chapter 2: Between Scylla and Charybdis: Sidgwick's Axioms of Ethics	29
Chapter 3: Universal Good	65
Chapter 4: 'Equal Good' and the Partial Dualism of Practical Reason	90
Chapter 5: Derivative Reasons and the Air of Paradox	119
References	143

Acknowledgments

Material from this dissertation was presented at the 13th Conference of the International Society for Utilitarian Studies in Yokohama, Japan, and the 8th Rocky Mountain Ethics Congress in Boulder, Colorado. I thank audiences at both events for helpful discussion. I am also grateful for the feedback I received from the members of the WUSTL Philosophy Department who attended my colloquium presentation in May 2015. For comments on all or part of the text, and/or for discussion, I would like to thank Nate Adams, Spencer Case, Jason Gardner, Jeff Dauer, Alex Dietz, David Phillips, Doug Portmore, Neil Sinhababu, Travis Timmerman, Martin Turner, David Winchell, and Peter Zuk.

Since the dissertation marks the culmination of a long educational journey, I would like to thank some of the people who helped me reach this point. I would like to thank my first philosophy professors – Jennifer Noonan and Michael Valle – for inspiring me to become a philosophy major through their brilliant teaching. I also thank my two advisors from my final years as an undergraduate – Chad Mohler and Stephen Pollard – for providing the support and encouragement I needed to complete my degree and apply for graduate study in philosophy.

I was fortunate to begin my graduate studies at the University of Missouri – St. Louis. I am forever indebted to Anna Alexandrova for instilling confidence and work ethic in me during my first semester of graduate school. I am also indebted to John Brunero and Eric Wiland for helping me develop as a writer and for continuing to support my professional development even after I completed the program at UMSL.

Washington University in St. Louis has been the ideal place to pursue a Ph.D. In addition to the top-level academic resources, I have also benefitted from generous financial support from

the Graduate School of Arts & Sciences and the Department of Philosophy. I am eternally grateful for the opportunity given to me by WUSTL.

I owe a substantial debt of gratitude to Christopher Heath Wellman for supporting me in numerous ways since I first set foot on campus in 2009. I am fortunate to have been working in the department during his tenure as chair. I thank him for embracing the mantle of leadership and for always having an open door.

Many thanks to Frank Lovett for serving on my dissertation committee, as well as for his general efforts toward promoting an atmosphere of collaboration between Political Science and Philosophy at WUSTL.

I thank Roger Crisp for agreeing to serve as the outside member of my committee. Roger's influence on my views will be obvious to anyone who reads the chapters that follow, and it is an honor to have his name appear on the title page.

Each of the four internal committee members has had a profound impact on me. I thank Julia Driver for being a wonderful advisor over the past six years. Her relentless pursuit of excellence as a philosopher has inspired me to push myself harder than I ever thought I could. I thank Charlie Kurth for always making time for me and my fellow graduate students no matter how much he had on his own plate. I thank Anne Margaret Baxley for reading countless drafts of my papers on Kant, and for expressing unwavering confidence in me. Lastly, I thank Eric Brown for providing a model to emulate as a father, teacher, philosopher, mentor, and friend.

I thank my family, and in particular my Mom, Dad, and Sister, for the love and support they have provided throughout my life, including the seemingly endless time I have spent as a graduate student.

A portion of this dissertation was written during a challenging time for me on a personal level. I would like to thank the following people who helped me keep my arms straight as I completed the project during the summer of 2015: Nich Baima, Kit Wellman, Eric Brown, Michelle Younger, Dustin Ireland, Nate Medeiros, Stacey Medeiros, Ryan Williams, Jeff Dauer, Elaine Dauer, Bryan Schlueter, Kat Schlueter, Bo Zierden, Kyle Dillingham, David McDonald, Erin Brugger, Kelly Brugger, Jesse Spina, Evan Judge, Sarah Judge, Dave Duley, Julie Barone Duley, Brian Hutchinson, Nicole Ernst, Julie Iromuanya, Sandra Tursini, Thomas Brown, Kéli Spring, Krystine Garcia, Ryan Benscoter, Jesselyn Benscoter, Slavka Jurinich, Ashlea Smith, Andrew McGinnis, Francesca Tomic, Mollee Huisinga, Kasey Leachman Rodgers, Asahi Oshiro, Claire Barwise, Bilal Khan, David Lewis, and Brian Whetter. Special thanks to my twelve amazing students at the JHU Center for Talented Youth who inspired me with their work ethic and philosophical acumen when inspiration was what I needed most: Jordan Bibby, Alison Cottingham, Dalinah De Lorenzo, Haoning Jiang, Andrew Kim, Skye Levine, Erin Liu, Joshua Park, Lucinda Quintal, Cinthia Sanchez, Brooke Schumaker, and Tiffany Zhao.

There are two people who have been especially pivotal in helping me complete the quest I began in the fall of 2006. Thank you Nich Baima for countless conversations about my work, for reading countless drafts, and for providing words of encouragement, sympathy, and wisdom through the trials and tribulations of graduate school...thick like blood...dogs for life.

Thank you Sarah Kertz for standing by me for ten years. It was your love and support, more than anything else, that enabled me to chase down my dream.

ABSTRACT OF THE DISSERTATION

Bound to Aim at Good Generally: A Sidgwickian Argument for Rational Impartialism

by

Tyler S. Paytas

Doctor of Philosophy in Philosophy

Washington University in St. Louis, 2015

Professor Julia Driver, Chair

In this dissertation I defend an unorthodox theory of practical reason called Rational Impartialism. According to this view, one ought to have equal fundamental concern for the equal good of everyone. Central to my argument for Rational Impartialism are two of Henry Sidgwick's axioms of ethics. According to the axiom of Equal Good, one ought to have equal fundamental concern for equal portions of the good. According to the axiom of Personal Irrelevance, the equal good of any two individuals is exactly similar qua portions of universal good. Sidgwick claims that from these two axioms we can deduce a principle of benevolence which is essentially the view I call Rational Impartialism. However, there are three problems with this deduction: (1) an egoist can accept the axioms but resist impartialism by insisting that 'good' is agent-relative, (2) if partiality towards loved ones is itself part of the good, then adherence to the axioms could be compatible with having ultimate reasons of partiality, (3) disagreement over the axiom of Equal Good seems to undermine its usefulness as a foundational ethical premise. I attempt to revive Sidgwick's derivation of Rational Impartialism by responding to all three challenges.

1 The Partiality Question

1 Sidgwick's Struggle

In his “great, drab book,” *The Methods of Ethics*, Henry Sidgwick is in search of fundamental ethical principles possessing clarity and precision akin to the foundational principles of geometry and the physical sciences.¹ After failing to find any such principles in the morality of common sense, Sidgwick offers a set of axioms which seem to him to possess the “highest certainty attainable” in philosophical ethics. One such axiom states that no individual’s good is of more importance from “the point of view of the universe” than the good of any other. A second states that rational beings are “bound to aim at good generally...not merely at a particular part of it.” From these two axioms Sidgwick deduces a principle of impartialism which holds that “each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him” (ME 382).²

Unfortunately, Sidgwick is unable to stand by this deduction as he believes there is a rival principle that is equally well justified. The rival principle is one of self-interest stating that “[one’s] own happiness is an end which it is irrational to sacrifice to any other” (ME 498). Because the principles of impartialism and self-interest regularly deliver conflicting verdicts about what an agent ought to do, Sidgwick somberly concludes that there is an irresolvable “Dualism of the Practical Reason,” meaning that in many circumstances, “practical reason, being

¹ The phrase “great, drab book” comes from Derek Parfit in the preface to *On What Matters* (2011: xxxiii).

² Henry Sidgwick, *The Methods of Ethics*, 7th ed. Indianapolis: Hackett, 1981 (reprint of London: Macmillan, 1907). I will refer to the 7th edition as ‘ME’. For earlier editions I will add the edition number (ME1, ME2, etc.). References to Sidgwick’s other writings on ethics and epistemology are to his *Essays on Ethics and Method*, ed. Marcus Singer (Oxford, 2000), abbreviated EEM.

divided against itself, would cease to be a motive on either side; the conflict would have to be decided by the comparative preponderance of one or other of two groups of non-rational impulses” (ME 508).

It should come as no surprise that Sidgwick’s careful, systematic, and ingenious study of the methods and principles of ethics culminates with the question of how to rationally adjudicate conflicts between self-interest and the general good. This is, in my view, the most fundamental and important question of philosophical ethics. It is also the most difficult. Most everyone agrees that we have basic reasons to promote the interests of all human beings as such. Yet, there are also strong intuitions suggesting that we have fundamental reasons to prioritize the needs of ourselves and our intimates. Hence, the unavoidable question: To what extent (if any) does one’s own good and the good of one’s loved ones carry greater rational weight for one than the equal good of strangers? I call this the *Partiality Question*. Though Sidgwick failed to see how we could be justified in settling on an answer to this question, my aim in this dissertation is to argue that there is a well-justified answer. I shall defend a controversial theory of practical reason called Rational Impartialism (RI). According to this view, one ought to have equal fundamental concern for the equal good of all individuals.³

Rational Impartialism is not a popular view.⁴ Although many philosophers believe the welfare of complete strangers is a source of objective, non-derivative practical reasons, few such

³ The ‘ought’ in this formulation is not the moral ‘ought’ but rather the ‘ought’ of ultimate (i.e. non-derivative) practical reasons). RI is compatible with an agent having derivative reasons to prioritize the good of some over the equal good of others. But this will only be the case if such partiality is ultimately justified from an impartial point of view. Note also that RI is formulated in terms of the ‘equal good’ of all individuals. This is partly to allow for the possibility that we are justified in having less fundamental concern for the well-being of agents with bad characters. If normatively significant desert exists (which is a question I shall not attempt to settle), then five minutes of happiness for Hitler and five minutes of happiness for Gandhi are not equally good, and so RI would not require equal fundamental concern for both.

⁴ It bears mentioning at the outset that Rational Impartialism is neutral with respect to consequentialist and deontological approaches to ethics. Consequentialists who believe there are prudential reasons that compete with impartial ‘moral’ reasons will reject RI. Deontologists who deny that there are duties and permissions of partiality

writers have defended impartialism, and several (perhaps most) have explicitly rejected it.⁵

While these writers agree that personal and partial interests are not the only source of practical reasons, they find complete impartialism untenable.⁶ Russ Shafer-Landau (2010: 110) claims that the principle that one's own interests carry extra weight "seems to be a basic, fundamental element in any plausible ethical view."

Though it is unconventional, I believe Rational Impartialism constitutes the most plausible answer to the Partiality Question. At the very least, the view merits more serious consideration than it has received to date. My strategy for defending RI is to return the axiomatic deduction of impartialism provided by Sidgwick. Sidgwick abandons his argument due to questions about the scope of 'good'. Had he been confident that each individual's good is part of a 'universal good', meaning that all rational beings have reasons to promote it regardless of how they are related to it, then he would have been confident that one ought to give equal weight to the equal good everyone. However, Sidgwick found it just as plausible that 'good' is always relative to the individual so that each agent has ultimate reason to pursue only her own good. In light of this worry, I shall attempt to vindicate the universal conception of 'good' about which Sidgwick was uncertain. I shall also argue that instantiations of partiality within loving relationships are only derivatively good. If I succeed on these fronts, we can add two further premises to Sidgwick's deduction: one stating that the good is universal, and another stating that partiality itself is not a part of universal good.

can accept RI. As I hope to make clear in the discussion below, RI should not be viewed as a moral theory but rather as a theory of ultimate practical reasons.

⁵ Crisp (2006: 145); Shafer-Landau (2010: 110); Parfit (2011, 1:131); Enoch (2011: 19). Enoch is more tentative in his denial of impartialism than the other authors cited.

⁶ Of course, philosophers such as Shelly Kagan (1989) and Peter Singer (1981) have famously defended impartialist views. But their arguments have typically been about the impartiality of morality rather than practical reason itself. A notable exception is the recent work of Singer and Katarzyna de Lazari-Radek (2014) which explicitly attempts to answer the question of what people ought to do in the 'decisive-reason' sense rather than the 'moral' sense.

While the vindication of these additional premises likely would have convinced Sidgwick to embrace his initial impartialist conclusion, a compelling case for Rational Impartialism requires more.⁷ Unfortunately, one of Sidgwick's key axioms is much more contentious than he realizes. Sidgwick is aware that common-sense moralists will reject the axiom stating that rational beings are bound to aim at good generally rather than particular parts of it. This is because common-sense holds that there are basic duties of partiality. But Sidgwick is undaunted by this disagreement as he provides arguments that he takes to undermine belief in such duties, as well as debunking explanations for the intuitions that support them. However, what Sidgwick could not have anticipated is that two of his greatest disciples—philosophers who share many of his methodological commitments as well as his ethical and metaethical views—deny that reason requires aiming at all of the good equally (despite their belief in universal good). Roger Crisp (2006) and Derek Parfit (2011) hold that while everyone's welfare is part of universal good, one's own good does in fact carry greater rational weight for one than the equal good of others. The explanation given for the extra weight of self-interest is that value is not the only source of reasons—it also matters how one is related to the value in question. In light of this dissent from two leading Sidgwickians, my argument for Rational Impartialism requires a defense of Sidgwick's axiom against the challenge presented by Parfit and Crisp.

The rest of this dissertation comprises chapters that outline the Sidgwickian argument in more detail and defend the controversial premises. The rest of the present chapter is devoted to laying the groundwork for investigation of the Partiality Question. In the next section I argue that while pursuing answers to fundamental ethical questions it is best to avoid concepts and terms that are distinctively 'moral'. In section 3 I expound on the concepts that are most

⁷ Sidgwick would not have needed the premise about the derivative value of partiality because he takes himself to have successfully defended a hedonistic value theory. Since I do not wish to defend hedonism here, my argument does require showing that instantiations of partiality are not intrinsically valuable.

important for present purposes including the notion of a practical reason. I also explain why the common tendency to focus on reasons that are relativized to different practical domains is both misguided and detrimental to philosophical progress. Section 4 comprises a discussion of Sidgwickian metaethics and some of the assumptions about the nature of normativity that I shall be proceeding under throughout the dissertation. A brief conclusion follows.

2 Taking Morality Less Seriously

Although philosophers have long thought that adjudicating the conflict between self-interest and the greater good is a matter of determining what morality requires of us, I believe focusing on distinctively moral concepts when investigating fundamental practical questions is a mistake. I shall now provide three general considerations that explain why I set aside questions of moral obligations and permissions for purposes of this dissertation.

2.1 Questionable Normative Significance

The first reason why I have chosen not to utilize moral concepts is that the normative significance of purported moral facts is dubious at best. While there is no universally accepted account of what it means to be under a moral obligation, there is a substantial amount of agreement that blameworthiness is central to the concept.⁸ When we fail to meet our moral obligations, we are said to have done something ‘morally wrong’. To say that an action is ‘morally wrong’, on one common conception, is to say that an agent who performs it is

⁸ See Skorupski (1999 ch. 7) and Crisp (2006 ch. 1).

blameworthy, which amounts to her meriting emotional sanctions in the form of resentment from others and feelings of guilt within herself.⁹

It has been plausibly suggested that in order for morality to function effectively, the blamed agent must judge that the ‘wrongness’ of her action was itself an ultimate reason not to have done it.¹⁰ But when we stop to reflect on whether moral wrongness is itself a reason not to perform an action, doubts emerge. Consider an act widely held to be morally wrong such as killing an innocent person for the sake of amusement. What are our reasons for refraining from performing such an act? One obvious candidate is the fact that killing a person deprives her of all the value of her future.¹¹ A second is that the friends and family of the victim will experience tremendous emotional suffering as a result. These considerations might explain why it would be wrong to perform the action, if we understand ‘wrong’ to mean that there are decisive reasons not to do it. But if we use the moral sense of ‘wrong’ we imply that killing has an additional, distinctively moral property which provides a *further* reason not to do it. But what are the grounds for believing that there is this further reason? The answer to this question is far from obvious.¹²

Perhaps the further reason provided by an action’s being morally wrong is grounded in facts about blameworthiness. Moral facts might have normative significance in that they tell us what agents are required to do in order to avoid being blameworthy. But why should we think that our ultimate practical reasons include reasons to feel the emotions that are central to the

⁹ Mill (1998); Strawson (1962); Gibbard (1990); Skorupski (1999); Prinz (2009). For discussion of alternative senses of ‘morally wrong’ see Parfit (2011, 1: 164-74).

¹⁰ Mackie (1977); Joyce (2001).

¹¹ Marquis (1989).

¹² Of course, many people do take moral wrongness to be a reason-giving feature of an action. But given the evolutionarily influenced structure of moral emotions and moral judgments, it is unsurprising that those of us who have been raised in societies that employ positive moralities would view moral wrongness in this way. See Crisp (2006: 16).

concept of blameworthiness? Take moral anger for example. It is true that human beings typically do feel anger toward those who lie, cheat, and steal. But what makes it the case that we have reasons to feel this way? It is doubtful that feeling anger is valuable for its own sake. Perhaps we should feel anger because this emotion motivates us to perform actions, such as punishing those who harm us, that may be beneficial as a means of deterrence. But notice that if this is right, our reason for feeling anger is merely derivative; we have a reason to feel anger only insofar as it will motivate us to perform blaming actions that we have reason to perform. Moreover, it is often the case that we do not have any reason to act as our moral emotions prompt us to, such as when expressing anger would do nothing but escalate a conflict.

Some theorists might be unmoved by these considerations because they believe moral facts tell us about which emotional responses would be ‘fitting’, and fittingness is thought by some to be a fundamental normative notion.¹³ One reason why fittingness has garnered attention recently is that this concept seems to track the so-called “right kind of reasons” to have certain beliefs and non-cognitive attitudes. For instance, the fact that some evil despot threatens to torture an innocent child unless I desire for that child to be tortured does not mean that I ought to have this desire. Whether I have a reason to desire some state of affairs seems to depend on the intrinsic features of that state of affairs rather than the consequences of having the desire. Thus, while I would have reason to try to cause myself to have this desire, if I did come to possess it I would not be in a fitting motivational state. Perhaps some might think that fittingness can also explain the normative relevance of moral emotions. If someone causes harm to an innocent person, it is plausible that it would be fitting for others to feel anger towards the perpetrator and for the perpetrator to feel guilt.

¹³ See D’Arms and Jacobson (2000); Chappell (2012); and McHugh and Way (forthcoming).

The problem with the appeal to fittingness is that it is unclear how facts about the fittingness of feelings could have fundamental ethical significance. As Sidgwick notes at the beginning of the *Methods*, ethics is an investigation into “what individual human beings ‘ought’—or what is ‘right’ for them—to *do*, or to seek to realize *by voluntary action*” (ME 1, emphasis added). Since our emotional responses are not under our direct voluntary control, the most we can do is try to preserve or change the feelings we have, and to try to cultivate the best possible psychologies. And when deciding whether to preserve or change an emotional response, the question of whether the response is fitting is irrelevant. To see this, suppose that you attend a professional basketball game, and your name is drawn for a chance to win a million dollars by making five consecutive free-throws during halftime. Given these circumstances, it would certainly be fitting for you to feel extremely nervous as you step up to the free-throw line. But suppose that, surprisingly, you feel cool, calm, and collected. Does this mean you have reasons to try to make yourself nervous by focusing on the magnitude of the stakes, just so that you can have the fitting response to your circumstances? Of course not. On the other hand, if you do start to feel nervous it seems clear that you would have decisive reason to try to calm yourself down so that you can sink the free-throws. This illustrates the point that when it comes to ultimate practical reasons, facts about which emotional responses would be fitting are unimportant. Thus, even if it is true that it would be fitting to feel anger towards someone who violates a moral requirement, this tells us nothing about our ultimate practical reasons.

2.2 First-Order Judgments

A second reason for shifting our focus away from moral concepts when engaging in philosophical ethics is that emphasizing moral concepts can have deleterious effects on our first-

order judgments. By now, most everyone working in ethics agrees that the mere fact that an ethical judgment is accepted by common-sense morality is not sufficient warrant for accepting it. This is because many of these judgements are predicated on distinctions that do not survive critical reflection. For instance, because common-sense morality attaches substantial weight to the distinction between doing and allowing, many otherwise conscientious people believe it is better to let someone die a slow and painful death rather than painlessly killing him. Another example is physical proximity. Given that throughout most of human history there was nothing a person could do to promote the interests of distant strangers, it is unsurprising that moral intuitions tend to place at least some weight on physical distance. When our normative capacities were developing, proximity mattered a great deal. But when we think carefully about whether mere physical proximity has ultimate significance for determining whether or not we should provide aid, we realize that intuitions that push us towards prioritizing those who happen to be closer to us are unreliable.¹⁴

Even though few philosophers accept the verdicts of common-sense morality without subjecting them to critical scrutiny, it is plausible that our use of moral concepts while engaged in ethical theorizing increases the likelihood that arbitrary distinctions which are vestiges of our evolutionary past will have undue influence. This is because describing an action as ‘morally wrong’ or ‘cruel’ is likely to elicit emotional responses that inhibit our ability to consider a wide range of factors.¹⁵ Even if philosophers are generally capable of overcoming these effects, it would still be wise to limit the risk. Thus, rather than asking what a given agent is morally obligated to do, I think we can do better by asking what a given agent has most reason to do. By

¹⁴ For additional examples and further discussion see Crisp (2006: 20-27).

¹⁵ Crisp (2006: 20-24).

de-emphasizing ‘morality’ we stand a better chance of being guided by our powers of reasoning.¹⁶

It bears mentioning that the prioritizing of reason over feelings that I am advocating does not imply that feelings have no role to play in the life of a successful agent. Given the facts about human nature, it is plausible that we have reasons to cultivate certain emotional dispositions in ourselves, such as the disposition to feel sympathy when others are suffering, and to feel anger towards those who wantonly cause harm. Thus, my position is not that feelings do not or should not have a role to play in an ethical life. The claim is rather that we should try to avoid being influenced by feelings when searching for fundamental ethical truths. One way of doing this is to focus less on morality and more on practical reasons.¹⁷

2.3 Metaethical Judgments

A third problem generated by focusing on morality is that doing so can lead us astray in our metaethical theorizing. The more one reflects on the origins of morality as well as the nature of purported moral properties, the more attractive moral anti-realism – the view that there are no response-independent¹⁸ moral truths – becomes. There are several reasons for this. First, as already mentioned, there seems to be little motivation for accepting the existence of moral properties with independent reason-giving force once we consider other possible sources of reasons not to perform “wrong” actions.

¹⁶ Cf. Norcross (2008). See also Crisp (2006: ch. 1).

¹⁷ I should also note that my reservations about using moral concepts while theorizing about ethics do not imply that moral concepts should be jettisoned from everyday thought and talk. It may be that these concepts are so central to our current way of life that trying to eliminate them would do more harm than good. See Crisp (2006: 27) and McElwee (2010)

¹⁸ I borrow the term ‘response-independence’ from David Enoch (2011: 3). He provides the following characterization of response-independence: “Whether or not a given normative statement applies (for instance) to a given action does not depend on what attitudes regarding it – cognitive or otherwise – are entertained by those judging that it is (or is not) or by anyone in their environment, nor does it depend on the attitudes, desires, and the like of the agent whose action it is or of anyone in *her* environment.”

Second, there is no denying that the moral concepts and the institution of morality they belong to have been heavily shaped by biological and social contingencies. In the case of biology, it is widely accepted that evolutionary forces played a significant role in the development of the ‘reactive-attitudes’ that are essential to moral discourse and practice.¹⁹ These forces also plausibly played a role in shaping some of the intuitions and judgments that compose much of common-sense morality.²⁰ Social and cultural evolution have also shaped our positive moralities. Morality is a social institution, and the views and practices of societies and cultures that arise largely from accidental circumstances play a major role in determining which views are held and which concepts gain footing. To take one notable example, Schopenhauer (1995) and Anscombe (1997) both argue that central moral concepts such as ‘forbidden’ and ‘permissible’ are vestiges of divine command ethics, and are only legitimate under the assumption that there is an external lawgiver. These biological and historical contingences cast serious doubt on the likelihood that our moral judgements correspond to response-independent normative truths. Moreover, because the tendency to believe in objective moral truths can be easily explained without having to posit any genuine moral properties, considerations of parsimony tell in favor of anti-realist or nihilist views of morality.

Once we become skeptical about the existence of response-independent *moral* truths, it is easy to arrive at the conclusion that there are no response-independent truths about what we have reason to do. This is partly due to the fact that so many of our metaethical inquiries have focused primarily on morality.²¹ But we must be careful not to let doubts about morality influence our views about normativity. While there are compelling reasons for thinking that morality is merely

¹⁹ Prinz (2009)

²⁰ Street (2015)

²¹ Scanlon (2014: 1) points out that up until the late 1970s, metaethical debates focused almost entirely on morality. Though today there is more discussion of reasons in general, morality still seems to garner the most attention.

a human construction, I do not believe the same is true in the case of normativity. Even if I am not genuinely under a “moral obligation” to promote the welfare of others, this would not preclude the possibility that the welfare of others can ground genuine response-independent reasons.

In addition to the reasons I have given for bracketing the use of moral concepts, there is also a basic point that pertains to the fundamentality of different practical questions we might ask. It seems clear that the question of what I have most reason to do is more basic than the question of what I am morally obligated to do. One way of seeing this is to note that even if there are moral obligations which are independent sources of reasons, we still must answer the question of how these reasons compare with other considerations such as self-interest. An answer to the question of what I am morally obligated to do would not be sufficient for settling the question of I have most reason to do.²²

Perhaps the belief in moral properties will ultimately be vindicated, and distinctively moral concepts will help us resolve our practical concerns. But for the reasons I have given, it seems that the best strategy is to begin our investigation by thinking in terms of reasons rather than moral obligations, duties, and permissions.

²² Although the *Methods* includes lengthy treatments of the morality of common sense, along with substantial discussions of distinctively moral concepts, Sidgwick does recognize that the most basic and fundamental questions in ethics concern reasons rather than moral obligations and duties. Consider the following remarks from the opening of Book I, chapter VI: “The aim of Ethics is to systematize and free from error the apparent cognitions that most men have of the rightness or reasonableness of conduct, whether the conduct be considered as right in itself, or as the means to some end commonly conceived as ultimately reasonable” (ME 77). Crisp (2015: 18) notes that the ‘or’ in this passage is expegetic, and hence that the reference to ‘rightness’ can be eliminated. Thus, Sidgwick’s point is that philosophical ethics is at bottom about our ultimate practical reasons. Here Crisp follows the reading offered by Jerome Schneewind in his influential book on Sidgwick. Schneewind claims that for Sidgwick, “[t]he point is to determine how we are to decide what is ultimately reasonable to do, regardless of whether the ultimately reasonable is called ‘moral’ or not” (1977: 228; reference from Crisp 2015: 18).

3 Reasons and Practical Domains

3.1 Practical Reasons

The Partiality Question is primarily about our practical reasons, which are a particular type of normative reason (the other type being epistemic). Thus, before we can consider the various candidate answers to the Partiality Question it will be necessary to clarify the concept of a normative reason and highlight some of the important distinctions pertaining to this concept. I share the common view that the notion of a normative reason cannot be helpfully explained via a definition.²³ It is sometimes said that a normative reason is a consideration that counts, for the agent, in favor of performing some action or having some belief or attitude.²⁴ But as Parfit (2011, 1: 1) notes, ‘counts in favor of’ is roughly synonymous with ‘gives a reason for’. Parfit suggests that because the notion of a normative reason is fundamental, the way to explain it is to get people to think thoughts that employ the concept and to explain its relation to other notions.²⁵ He offers the example of the thought that we always have a reason to want to avoid being in agony (ibid.). Another example is the thought that the fact that a neon ‘OPEN’ sign is glowing in the window is a reason to believe the store is open.

In addition to getting people to think the right thoughts, it is also useful to contrast the notion of a normative reason with other concepts involving the word ‘reason’. In particular, it is important to contrast normative reasons with explanatory and motivating reasons. An explanatory reason is, as the name implies, a fact that explains why some particular event

²³ Parfit (2011, 1:1); Raz (2011: 18).

²⁴ In including the ‘for the agent’ clause I follow Crisp (2006: 38). The purpose of this clause is to allow for properties that might count *for some other agent* in favor of performing an action or having some belief even though those properties are not reasons for the agent in question. To borrow Crisp’s example, if egoism is true, then the fact that an action of yours would benefit me counts *for me* in favor of your performing it, though this would not give you an ultimate reason to perform it (ibid. n. 4).

²⁵ Sidgwick makes a suggestion along these lines concerning the concepts of ‘ought’ and ‘right’ (ME 32-3).

occurs.²⁶ For instance, if a car does not start due to battery failure we might say that the fact that the battery is dead is the reason why the car will not start. In a certain sense, normative reasons are a type of explanatory reason. This is true insofar as facts about what one has reasons to do or believe can always be part of an explanation of something. However, normative reasons are distinct from explanatory reasons in that in addition to explaining, they also *favor* certain actions and attitudes. Moreover, normative reasons do not merely explain why an action or attitude is favored—they occupy a favoring relation that is prior to and independent of explanation. As Joseph Raz (2011: 19) puts the point: “Their explanatory use is secondary, and depends on the fact that they favor what they favor, a fact that sets them apart from other explanatory reasons. The existence of a normative relation: that one thing is a reason for another, is, on this suggestion, the object of the explanation.”

Motivating reasons are a type of explanatory reason. They are facts that explain why an agent performs a particular action. When I drive to the grocery store, my motivating reasons might be my desire for food, and my belief that I can obtain food at the grocery store. Importantly, an agent’s motivating and normative reasons often diverge. For instance, though a sadist’s motivating reason for torturing his victim is his desire to witness suffering, we could plausibly say that he has no normative reason to torture. And even if we think that the enjoyment he will derive from inflicting pain is a normative reason for him, we could still claim that he has decisive reason to refrain from doing so despite his motivational state.

Once we turn our focus specifically to normative reasons there are still several distinctions we can draw. The philosophical literature on reasons is replete with dozens of distinctions and various terminological conventions. For present purposes it is only necessary to

²⁶I follow Parfit in using ‘event’ in a wide sense which includes both acts and states of affairs.

discuss a few such distinctions. The first is that between epistemic and practical reasons. Whereas epistemic reasons are reasons for having a certain belief, practical reasons are reasons for performing a certain action. Since ethics is primarily concerned with conduct rather than beliefs, we shall focus primarily on practical reasons.

One of the most useful distinctions pertaining to practical reasons is that between *ultimate* reasons and *derivative* reasons. An ultimate reason is a reason to perform an action that cannot be explained by any further reason. In contrast, a derivative reason is a reason to perform an action that is explained by some more fundamental reason. To illustrate, consider an agent who purchases a ticket to a concert. It is plausible that this agent had a reason to purchase the ticket because he will enjoy the concert, and he has a reason to seek enjoyment. Thus, we can say that his reason to purchase the ticket derives from his reason to seek enjoyment. But we might have trouble finding a reason why he should seek enjoyment. Enjoyment is thus a plausible example of a source of ultimate (or non-derivative) reasons. The reason to seek enjoyment is an ultimate reason if it does not inherit its normative force from any other reasons.²⁷

3.2 Practical Domains

A final distinction that is important for present purposes is not between two different types of practical reason but rather between genuine practical reasons and what I shall call *domain-relative* reasons.²⁸ We often speak of reasons an agent has relative to a particular practical domain such as fashion, etiquette, law, football, chess, etc. For instance, we might say that even if the interests of non-human animals give us strong reasons to refrain from consuming animal

²⁷ For further discussion see Crisp (2006: 8).

²⁸ I borrow this term from Stroud (1998: 172).

products, we nonetheless have reasons of fashion to wear leather shoes and mink coats. Reasons we are said to have relative to the standards of fashion are an example of domain-relative reasons (or *d*-reasons).

I believe it is a mistake to assume that our ability to think in terms of various practical domains and our corresponding tendency to refer to reasons that are relative to those domains is evidence that different domains are independently normative.²⁹ Of course we can make intelligible claims about “moral reasons,” “fashion reasons,” and perhaps even “Blue Öyster Cult fandom reasons.” We might say that I have reasons of fashion and reasons of BÖC fandom to wear my vintage “Black & Blue Tour” t-shirt.³⁰ And suppose it is true that the t-shirt is fashionable and my wearing it would express support for the greatest sci-fi-gothic classic rock band of all-time. These facts would not be enough to settle the question of whether there is anything counting in favor of my wearing the t-shirt. We could not know whether compliance with fashion or BÖC fandom norms is a reason to Φ without answering the more fundamental question of whether we have any reason to care about these respective domains. As Parfit notes, in order to make significant claims about the relative importance of the requirements of different domains, “we need some impartial, neutral criterion. Reasons provide such a criterion” (2011, 1: 148).

Do we have reasons to care about fashion and BÖC fandom? Perhaps. But if we do, it would seem that this is so only insofar as complying with these norms can make one’s life or the lives of others better, or if such compliance would instantiate a type of human excellence.³¹ If complying with these norms would not have these positive results, then it is hard to see what

²⁹ This point is raised by McLeod (2001: 274-5) and Dorsey (2013: 131-6).

³⁰ In 1980 Blue Öyster Cult and Black Sabbath co-headlined a highly successful nationwide tour.

³¹ Although I am tempted by welfarism about reasons, I shall not pursue this issue here. For a compelling defense of welfarism about reasons see Crisp (2006: ch. 2).

could count in favor of doing so. Thus, any normativity inhering in these practical domains could only be derived from things that matter at the fundamental level. Hence “reasons” of fashion and B.Ö.C. fandom are not genuine normative reasons.

Here one might object that we need justification for accepting the claim that reasons have fundamental normative importance such that they can undermine the verdicts of a given practical domain. If fashion norms tell me to Φ , why must I ask the further question about whether I have reasons to care about fashion? And if the norms of fashion cannot themselves vindicate the importance of fashion, how can reasons themselves vindicate the importance of reasons? Parfit considers this objection and provides the following response:

As this objection rightly claims, we cannot show that reasons matter by appealing to claims about reasons. But justifications must end somewhere. And if reasons are fundamental, we should not expect that we could justify the reason-involving criterion of importance, by appealing to some other, deeper criterion. Reasons *are*, I believe, fundamental. Something matters only if we or others have some reason to care about this thing. (2011, 1: 148)

Someone who thinks she has sufficient reason to wear fashionable clothes even though this will result in great suffering need not be conceptually confused. I have not been arguing that the claim that fashion is more important than things like welfare is incoherent. What I have been arguing is that an action or attitude being recommended by some practical domain or other does not by itself count in favor of performing that action or having that attitude. Practical domains by themselves are normatively inert. The verdicts of practical domains can only matter insofar as we have reasons to care about them. Hence, rather than saying that reasons are the fundamental normative standpoint, it might be more helpful to say that reasons are the *only* normative standpoint.

At this point one might wonder about two particular practical domains that seem most likely to have independent normative significance—prudence and morality. Though it is initially very plausible that these domains are genuinely normative, I believe this as an illusion. My primary aim in writing this dissertation is to make a convincing case that the fact that Φ -ing is in my self-interest is not an independent reason for me to Φ over and above the fact that Φ -ing promotes the interests of *someone*. Of course, if Φ -ing would promote my welfare this does indeed count in favor of my Φ -ing. But this reason is not grounded in my relationship to myself but rather the fact that my interests matter just as the interests of everyone else do. Thus, on the view I am defending it is false to say that prudence is an independently normative domain.³²

In section 2 I provided several reasons for turning our focus away from morality while investigating the Partiality Question. Among the reasons I gave were (1) the fact that settling the question of what I morally ought to do does not settle the question of what I ought to do, (2) skepticism about the importance of the emotional responses that are closely tied to fundamental moral concepts, (3) the distorting influence of moral concepts on our first-order practical judgments, and (4) the fact that the plausibility of anti-realist views of morality erroneously leads many people to accept anti-realism about practical normativity. As we saw, most of my skepticism about morality arises from concerns about the conceptual ties to emotional responses. However, there are conceptions of morality that are not in any way tied to emotions. One such conception holds that moral reasons are simply other-regarding reasons. Given my theoretical commitments it might seem as though I should accept the verdicts of morality thus understood as genuinely normative. But in fact, I deny that morality is independently normative even when it is conceived as simply a matter of other-regarding reasons. My reasons for denying normativity

³² Denying the independent normative significance of self-interest obviously requires substantial argument. My argument for this claim is presented in the chapters that follow.

here are analogous to my reasons for denying the normativity of prudence. While I obviously think that the fact that Φ -ing would promote the interests of many people is a reason to Φ , the fact that the people affected are *others* is irrelevant. On the view I am defending, I have equal reasons to promote equal interests regardless of how I am related to them.³³

Suppose for instance that I am faced with two options for how to distribute benefits between myself and some other agent. Option 1 would give me a benefit of 1, and the other person a benefit of 2. Option 2 would give me a benefit of 2 and the other person a benefit of 1. To keep things simple we can suppose that nobody else will be affected by the distribution. On the conception of morality we are considering I would have decisive moral reason to choose option 1. This is because option 1 is best for others (in this case the single other agent affected by my choice). But I believe the fact that option 1 would be best for agents other than me is irrelevant. What ultimately matters is the interests at stake regardless of how I am related to them. Thus, I would have equal reason to choose either option. Hence, on the view I am defending, morality is not independently normative even when conceived of simply as a matter of other-regarding reasons.

4 The Objectivity of Ethics

It is sometimes thought that answers to first-order ethical questions such as the Partiality Question are completely independent from answers to metaethical questions about the nature of practical reasons. One of the points that I hope to make clear in the chapters that follow is that it is a mistake to assume that ethics and metaethics are completely autonomous in this way. On the

³³ Some might object to Rational Impartialism not because it fails to give enough weight to self-interest but rather because it gives too much. The worry is that we tend to think an agent can have sufficient reason to sacrifice some of her interests for the sake of the lesser interests of others (e.g. sacrificing my life to save a stranger's leg). I address this objection in chapter 5.

contrary, I shall argue that facts about the fundamental sources and nature of practical reasons bear strongly on the question of whether there are ultimate reasons of partiality. More specifically, I will suggest that if practical reasons are grounded in objective facts about actions, objects, and outcomes, rather than in facts about our desires and evaluative attitudes, this tells in favor of an impartialist account of practical reason. My aim in the rest of this section is to explain why I shall be proceeding under the assumption that practical reasons are indeed grounded in objective facts about acts, objects, and outcomes that make them worth performing, promoting, or producing for their own sake.

4.1 Two Kinds of Theory

Suppose a friend tells you that you have a reason to visit the science museum this weekend. And suppose further that your initial response is rather skeptical: “What reason do I have to go to the science museum? Aren’t there better things for me to do with my time?” At this point, your friend will likely try to explain why it is true that you have a reason to visit the museum. Perhaps she will point out that they are opening a new exhibit on a topic that she believes you will greatly enjoy learning about. Suppose your friend is right, and you would in fact enjoy the new exhibit. If so, most of us would agree that your friend was also right about your having a reason to visit the science museum. Perhaps the reason is not decisive because there may be even better things for you to do, but you seem to have at least some reason nonetheless.

While most philosophers would accept that you would have some reason to visit the science museum in the above example, there is great controversy about what would make this true. According to some theorists, the ultimate explanation for the existence of a reason to perform an action such as visiting the museum could only be that doing so would satisfy one or

more of your desires. These writers defend subjectivism about practical reasons. Though subjectivism comes in many forms, what all versions share in common is the idea that whenever an agent A has a reason to Φ this fact is grounded or explained by the fact that Φ -ing would result in the satisfaction of one or more of A 's desires (or other subjective concerns such as valuing or preferences). Most (if not all) proponents of subjectivism accept that it is not our actual subjective concerns that ground our reasons but rather the subjective concerns we would have if we were made aware of all the relevant facts.

According to an alternative family of views, your reason to visit the science museum is not grounded in the fact that doing so would accord with your valuing or satisfy your desires. Rather, your reason is grounded in the objective features of either the museum itself or the states of affairs that would result from your visit (e.g. the enjoyment you would experience).

According to these views, our practical reasons are object-given, meaning that they are provided by facts about the features of objects, acts, and outcomes that make them worth performing or promoting for their own sake. These reason-giving facts usually make these acts, objects, and outcomes good or bad for particular individuals, or impersonally good or bad. Hence, we can also refer to object-given theories as value-based (Parfit 2011, 1: 45). According to these theories, an agent can have a reason to Φ even if Φ -ing would not satisfy any of her actual or idealized subjective concerns.

There are several theoretical motivations for accepting a subjective rather than an objective account of practical reasons. Among the chief motivations is a commitment to metaphysical naturalism. One result of the impressive accomplishments of modern science is that philosophers have become increasingly wary of philosophical views that appear at odds with the modern scientific worldview. Objective theories are a prime example of such views.

Objective value and reasons are not the sorts of things that can be discovered or verified by any of the natural or social sciences. While science cannot conclusively rule out the existence of non-natural entities such as object-given reasons, scientists do not appeal to such entities in their explanations of how the universe works. Because the natural and social sciences make no mention of objective reasons and value, many philosophers find belief in these things no less dubious than belief in ghosts and goblins. It is believed that the only entities that have a place in a respectable ontology are the purely ‘natural’ entities that are part of the best scientific theories. Moreover, even if there is objective value in the universe, many theorists are puzzled about how we could have justified beliefs about this value. If there are truths about objective value, they do not appear to be analytic, and it is unlikely that they are causally efficacious. It is thus difficult to explain how we could possess knowledge of such truths.³⁴

Of course, even philosophers who reject objective theories recognize that ordinary human discourse is committed to the existence of reasons and value. We regularly describe objects as good for people, actions as morally wrong, decisions as justifiable, choices as important, etc. Those who are committed to metaphysical naturalism can avoid the strikingly revisionary claim that all normative discourse is hopelessly defective by adopting a subjectivist account. Subjectivism holds that many of our claims about reasons and justification are true, and that this can be explained without running afoul of naturalism. What makes normative claims true, according to subjectivists, is that some specified class of (perhaps idealized) agents adopts a particular attitude toward the relevant object, action, decision, or choice. Value exists in the world because *we* value things. Things matter in the sense that they matter *to us*. Further, there

³⁴ The epistemic challenges facing non-naturalist versions of normative objectivism take a variety of forms. One of the most influential is Sharon Street’s “A Darwinian Dilemma for Realist Theories of Value” (2006). Some compelling attempts to meet this challenge can be found in Skarsaune (2011) and de Lazari-Radek and Singer (2014).

is no mystery about how we can come to have justified beliefs about our reasons if subjectivism is true. All we need to do is figure out which available action is most consistent with our desires and/or evaluative commitments.

A second motivation for subjectivism is the intuition that practical reasons are the sorts of things that should be capable of motivating agents. According to the popular Humean theory of motivation, an agent can be motivated to Φ only if she believes Φ -ing would satisfy one or more of her desires. Since motivation requires the presence of a desire, and reasons are presumably the sort of things that we can be motivated by, it would seem that we can only have a reason to perform some act if doing so would satisfy one or more of our desires. While subjective views are able to accommodate this thought, objective theories cannot. If practical reasons are object-given and value-based, then an agent might have a reason to Φ even if Φ -ing is in no way connected to the things she cares about. And if the Humean account of motivation is true, this would imply further that an agent might have a reason to Φ even if she could not possibly become motivated to Φ via rational means.

4.2 Sidgwick and the Presuppositions of Ethical Inquiry

There is a vast literature on the aforementioned attractions of subjectivism. Proponents of objective theories argue that the purported attractions are much less compelling upon reflection.³⁵ Subjectivists respond by sharpening their arguments and articulating further considerations in favor of their view.³⁶ For present purposes, it will not be necessary to recount the details of the various moves that have been made in this debate. This is because, as I shall now argue, regardless of the relative theoretical merits of the competing views and the arguments

³⁵ Crisp 2006; Parfit 2011; Scanlon 2014; Enoch 2011

³⁶ Street (2015); Sobel (unpublished manuscript)

meant to support them, the task I am undertaking in this dissertation requires that we proceed under the assumption that subjectivism is false, and that practical reasons are object-given and value-based.

In the first chapter of the first edition of the *Methods*, Sidgwick claims that the objectivity of ethics is a presupposition of the ethical investigations of philosophers and laypersons alike:

But if science of [ethics] be possible it must in another sense have objective existence: it must be an object of knowledge and as such the same for all minds. Even this however is not always admitted. “What ought to be done” (it is said) “is merely what *we* should like to be done, or, more precisely, what excites in us a specific feeling called approbation, varying in its object from mind to mind... But if it be maintained that two men may act in two different ways under circumstances precisely similar, and yet neither be wrong because he thinks himself right: then the common notion of morality must be rejected as chimera. That there is in any given circumstances some one thing that ought to be done and that this can be known, is a fundamental assumption, made not by philosophers only, but by all men who perform any process of moral reasoning. (ME1 5-6)³⁷

The final sentence of this passage is especially telling. Sidgwick’s point is that the common practice of engaging in a process of reasoning to try to discover what we ought to do is predicated on the thought that the answer to this question exists independently of our subjective attitudes and feelings. The Partiality Question is precisely the sort of question that requires the type of reasoning Sidgwick refers to, and hence, investigation of it can only be sensible on the assumption that there are objective facts about our practical reasons.³⁸

To see this point more clearly, it will help to consider a vivid example of a practical conflict between self-interest and the general good. Every fall I spend part of my Sunday afternoons in front of the television watching football. In order to ensure that I can see my

³⁷This passage is quoted in Phillips (2011: 13).

³⁸ Admittedly, Sidgwick here refers to ‘moral’ reasoning rather than practical reasoning generally. However, we can assume that the point would generalize for Sidgwick because, as Crisp notes, the normative questions Sidgwick is primarily interested in concern ultimate reasons rather than distinctively moral properties (Crisp 2015: 18, 106).

favorite team play, I pay extra to subscribe to a satellite service that broadcasts all the games. As someone who teaches and writes about ethics for a living, it is hard not to think about whether my football watching habit is justified. Given the severe global poverty crisis that is taking place, wouldn't my disposable income do a lot more good if I sent it to Oxfam rather than DirecTV? This is a real-life case in which I must arrive at a decision about how to weigh my own interests against the interests of distant strangers.

Such first-person experiences are part of the motivation for a theoretical investigation of the Partiality Question. Our hope is that if we think carefully enough, and consider all of the arguments, we shall ultimately arrive at the truth about the existence or non-existence of ultimate reasons of partiality. We expect that this endeavor will require rigorous thinking and deep reflection. And we are open to the possibility that by prioritizing our own welfare as much as we do, we are making a grave mistake. Subjectivists believe they can accommodate the intuition that such normative mistakes are possible. For instance, on Sharon Street's (2012) version of subjectivism, such mistakes are possible insofar as some of our choices might do worse by the lights of our overall set of evaluative commitments (even if one or more of our desires is satisfied). So when I subscribe to satellite television rather than donating to famine relief, I would be making a mistake if making the donation would be more consistent with my overall (contingently arrived at) set of evaluative commitments. But surely this is not the kind of mistake the fear of which motivates our engaging in substantial ethical inquiry. When I am attempting to determine whether I have most reason to write a check to Oxfam or DirecTV, I am not hoping to discover whether I happen to care more about following my favorite team or saving the life of a starving child on another continent. I am trying to figure out which of these I *should* care more about. But according to proponents of subjectivism, none of us have any

reason to care about the things we do because our reasons are determined by what we already happen to care about.

In an important sense then, subjectivism is close to nihilism (Parfit 2011, 107). If things only matter insofar as they matter *to us*, we do not need to do philosophy in order to answer the Partiality Question.³⁹ To figure out how much deliberative weight to give the interests of others we need only to think about how much we care about their interests compared to our own. And even if we end up failing to act in accordance with our overall evaluative commitments, there could not be any reason to regret this fact because these commitments were not tracking any deep normative truths anyway.

The Partiality Question is worth pursuing only on the assumption that there are objective truths about our practical reasons. The question presupposes that even if all of an agent's deepest desires (whether actual or idealized) are satisfied throughout her life, there is a possibility that she has made grave normative mistakes nonetheless (perhaps by not giving sufficient weight to the interests of others). Subjectivism precludes this possibility altogether. In considering theories of rational weight, we endeavor to discover whether our tendency to favor ourselves is justified by the light of reason. If there is no reason independent of our own subjective attitudes this endeavor would be futile. As Parfit puts the point, "If Subjectivism is true, we would have to make our choices in the dark" (2011, 1: 46).

It is worth noting the further point that if subjectivism is true, even the most extreme degree of partiality towards oneself could be justified so long as the agent's evaluative commitments are skewed towards her own welfare to a great enough degree. And even worse, fundamental partiality towards members of one's race and gender could be justified in the same

³⁹ Sharon Street acknowledged this point during personal communication on May 1st, 2014.

way. This is deeply implausible. Members of the Ku Klux Klan are not rationally justified in prioritizing the interests of White Americans, even if this favoring is entailed by their desires and evaluative commitments (as it surely is). This is a further reason for setting aside the possibility that our reasons are ultimately grounded in facts about what we happen to care about.

I shall thus proceed under the assumption that our practical reasons are object-given and value-based, and that in order to answer to the Partiality Question we must look beyond our contingently given desires and evaluative commitments. This is not to deny that objective theories of practical reason face serious challenges. The metaphysical and epistemological objections canvassed above are far from trivial. But it is beyond the scope of this project to address those worries. The question I am investigating is how we are to rationally weigh the interests of others against our own. It may turn out that normativity is an illusion, and that at bottom we have no reason to try to answer the Partiality Question because we have no object-given reasons to do anything. But if this turns out to be the case, it could not be true that we had reasons *not* to try to answer it. On the other hand, if normativity is not an illusion, we most certainly have strong reasons to figure out how the interests of others are to be weighed against our own.

5 Conclusion

In this chapter I have introduced the question that this dissertation centers on, and I have provided a general outline of my strategy for answering it. I have also explained why I shall not be focusing on distinctively moral concepts, and I have clarified the concepts that will play a vital role in the chapters that follow. Lastly, I explained why I shall be proceeding under the assumption that practical reasons are object-given and value-based, as opposed to being

grounded in facts about our contingently given desires, preferences, and valuings, as subjectivists claim.

I began this chapter by recounting how Sidgwick's attempt to arrive at foundational ethical principles ended with the somber conclusion that practical reason is divided against itself. In the next chapter I return to Sidgwick in order to lay the groundwork for my Sidgwickian argument for Rational Impartialism. I shall delve into Sidgwick's ethical epistemology and the self-evident axioms to which it leads him. I shall also offer novel interpretations of the key axioms and explain how they will function as premises in my argument for RI.

2 Between Scylla and Charybdis: Sidgwick's Axioms of Ethics

1 Introduction

Sidgwick's approach to ethics is guided by the conviction that our inquiries will only lead to trustworthy conclusions if our reasoning begins with premises of the "highest certainty attainable." Such premises must have the appearance of self-evidence, meaning that they are free of arbitrariness, and at least some degree of justification is obtained merely by contemplating them (ME 293 f. 1; EEM 29). But this *prima facie* justification is not sufficient for highest certainty; the foundational premises must also survive critical scrutiny. We must make sure that the appearance of self-evidence is not an illusion arising from convention or familiarity. And we must check to see that the proposition and its implications conflict neither with the beliefs of other equally competent judges, nor with other principles we take to be self-evident (ME 338-42). Only if a proposition meets these criteria can it serve as a genuine axiom of ethics. Sidgwick's hope is that such axioms, if they can be found, will provide a secure foundation for ethics analogous to the foundation provided by the basic axioms of geometry.

After investigating the core principles of common-sense morality, Sidgwick concludes that the appearance of self-evidence evaporates upon reflection. Once the vagueness of these principles is removed, we see that many of them are beset by arbitrariness and inconsistencies, while others are merely concealed tautologies. However, Sidgwick does believe that careful reflection on ethical concepts and the nature of rational agency gives rise to a set of foundational propositions that meet the criteria for highest certainty. Among them are an axiom stating that one's own welfare does not constitute a greater portion of the good than anyone else's, and an

axiom stating that reason requires aiming at good generally, not just particular parts of it (ME 382).

Sidgwick uses these axioms to deduce a principle of impartialism stating that “each one is morally bound to regard the good of any other individual as much as his own...” (ME 382). However, much to his dismay, Sidgwick ultimately concludes that the axioms cannot provide trustworthy ethical guidance after all. This is due to the fact that one of the key axioms only has real significance on the assumption that ‘good’ is a universal end, meaning that it is something to be pursued by all rational beings *as such*—regardless of how they are related to it. Proponents of egoism challenge this assumption by arguing that good is always relative to the individual. Because he sees no way to demonstrate that the Egoist is wrong about this, Sidgwick concludes that egoism is as well-justified as impartialism. This leads to his belief that there is an irresolvable “Dualism of the Practical Reason” that reduces the “Cosmos of Duty” to a state of chaos (ME1 473).

Given my aim of defending a purely impartial account of ultimate practical reasons, it should come as no surprise that I believe Sidgwick’s conclusion is mistaken. Though he failed to realize it, much of the material necessary for arriving at a principled resolution of conflicts between self-interest and the greater good can be found within the *Methods*. I shall argue that when the two axioms mentioned above are combined with a few additional premises, we obtain a powerful argument for Rational Impartialism—the view that there are no ultimate reasons of partiality.

The rest of this chapter is divided into three parts. I begin in section 2 by explaining the motivations and advantages of Sidgwick’s moral epistemology. Next, in section 3, I offer novel interpretations of Sidgwick’s axioms and defend my formulations against various objections.

Finally, in section 4, I delve deeper into Sidgwick's infamous Dualism, and I explain how I shall try to avoid this bleak conclusion by supplementing the axioms with two additional premises.

2 Philosophical Intuitionism

Before he ever set out to write the *Methods*, Sidgwick was a devotee of the utilitarianism of Mill. What he found attractive about Mill's position was that it appeared free from the arbitrariness and dogmatism that infected moral systems based on absolute duties ascertained by intuition. In Sidgwick's words: "I found in [Mill's utilitarianism] relief from the apparently external and arbitrary pressure of moral rules which I had been educated to obey, and which presented themselves to me as to some extent doubtful and confused; and sometimes, even when clear, as merely dogmatic, unreasoned, incoherent" (ME xvii). Sidgwick labels the use of intuition for the attainment of absolute duties the "Intuitional Method" (C.D. Broad (1930) would later provide the now widely used label "deontological"). From the time he was an undergraduate at Cambridge, Sidgwick felt that the Intuitional method was beset by a lack of a unifying explanation for the various purported duties, as well as a lack of guidance for resolving conflicts of duty (ME xvii).

But it was not long before Sidgwick grew dissatisfied with Mill's defense of utilitarianism. Though both elements of the view—psychological hedonism and ethical hedonism—seemed plausible when viewed in isolation, Sidgwick soon recognized the deep discrepancy between the natural impulse to promote one's own happiness and the ethical end of promoting the general happiness. In Mill, Sidgwick was unable to find a satisfying explanation for why it would be rational to sacrifice one's own happiness for the greater happiness of others.

He notes particular dissatisfaction with Mill's suggestion that such sacrifice would be "heroic."

Sidgwick's remarks on this point merit a full recounting:

It was no use to say that if I was a moral hero I should have formed a habit of willing actions beneficial to others which would remain in force, even with my own pleasure in the other scale. I knew that at any rate I was not the kind of moral hero who does this without reason; from *blind* habit. Nor did I even wish to be that kind of hero: for it seemed to me that that kind of hero, however admirable, was certainly not a philosopher. I must somehow *see* that it was right for me to sacrifice my happiness for the good of the whole of which I am part. (ME xviii)

The recognition of a need to "see" that self-sacrifice was right marks a turning point for Sidgwick. It is this need that prompts him to re-examine the intuitional method in hopes of finding a fundamental intuition that might provide a more secure philosophical foundation for ethics.

Though Sidgwick comes to accept that justification in ethics is not possible without relying on intuitions, one of his primary aims in the *Methods* is to draw crucial distinctions among different ways in which intuitions are utilized in ethical inquiry. He distinguishes three different versions of intuitionism which he labels 'Perceptual', 'Dogmatic', and 'Philosophical'. Perceptual Intuitionism (henceforth PI) is the method of determining whether a particular action is right through an immediate judgement of conscience. For instance, when an individual has the opportunity to steal the wallet of an unsuspecting co-passenger onboard a train, she can immediately sense that the action would be wrong without the need for abstract ethical principles or rules. On this approach, one's conscience is a reliable guide to truths about the rightness or reasonableness of conduct in any given circumstance.

Sidgwick outlines several problems with PI. First, the judgments of our conscience tend to be inconsistent across time: "The same conduct will wear a different moral aspect at one time

from that which it wore at another, although our knowledge of its circumstances and conditions is not materially changed” (ME 100). Second, there is also wide variance in intuitions about specific cases between apparent epistemic peers. These intra and inter-personal conflicts give rise to serious doubts about the reliability of immediate judgments of conscience. Moreover, these discrepancies are well-explained by a third problematic feature of PI. Intuitive judgements about particular acts are vulnerable to the influence of our motivational psychology, such as our instinctual drive for self-preservation, which we should not expect to track deep ethical truths. As a result of our socio-biological history we experience many impulses and immediate evaluative judgments that reveal themselves as unjustifiable upon further reflection. A few ubiquitous examples are the tendency to give undue weight to physical proximity when considering reasons of beneficence, and the tendency to make different moral judgments about the same conduct depending on how similar the agent in question is to oneself.¹ These ‘gut reactions’ are often experienced as accurate perceptions of ethical truths. In some cases these feelings may lead us in the right direction, but whether they do is largely a matter of luck. This is why it is dangerous to rely solely on immediate judgements about particular acts.

Dogmatic Intuitionism (henceforth DI) is a method of justification based on the assumption that the rules of common-sense morality provide generally reliable ethical guidance, and that the ethicist’s job is merely to refine and systematize them in order to make the rules clear and precise. One of Sidgwick’s primary misgivings about DI is that there are circumstances in which different principles of common-sense morality recommend conflicting courses of action. Sidgwick fails to see how there could be a non-arbitrary way of resolving such conflicts other than appealing to utilitarian considerations. A further worry is that even if the

¹ See Crisp (2006: ch. 1), and Haidt (2001).

principles and duties of common-sense morality were consistent with each other, there does not seem to be a deep unifying explanation for them. In Sidgwick's words:

Even granting that these rules can be so defined as perfectly to fit together and cover the whole field of human conduct, without coming into conflict and without leaving any practical questions unanswered,--still the resulting code seems an accidental aggregate of precepts, which stands in need of some rational synthesis. In short, without being disposed to deny that conduct commonly judged to be right is so, we may yet require some deeper explanation *why* it is so. (ME 102)

The lack of explanation should decrease our level of confidence in the verdicts of this method. Even if we tend to agree with the majority of common-sense moral rules, the absence of a unifying explanation and justification raises doubts about whether they have a truly rational basis.

As Crisp (2015: 104) points out, proponents of DI might object that common-sense morality contains deeper explanations than Sidgwick suggests: "Promise-breaking, for example, can be seen to be wrong in itself, or wrong because it is unjust; helping others is good in itself, or good because it is benevolent; and so on." But as Crisp notes, Sidgwick would still find such a theory too unsystematic to warrant strong confidence. In Sidgwick's view, what we need is "one or more principles more absolutely and undeniably true and evident, from which the current rules might be deduced, either just as they are commonly received or with slight modifications and rectifications" (ME 102).²

The dissatisfaction with Perceptual and Dogmatic Intuitionism leads to Sidgwick's preferred method of justification in ethics which he calls 'Philosophical Intuitionism' (henceforth PHI). This method is characterized by the non-inferential grasping of abstract, self-evident

² Cf. Singer and De Lazari-Radek (2014: 67). For further discussion of the different versions of intuitionism see Crisp (2002; 2015) and Phillips (2011: ch.5).

propositions that can provide a secure basis for ethical theories and principles. There are thus three key features of PHI that distinguish it from its rivals: abstraction, foundationalism, and self-evidence.

With regards to abstraction, the plausible suggestion is that by considering ethical propositions in the abstract rather than relying on our ‘gut reactions’ to particular sets of circumstances we are less likely to have our judgments distorted by the biases and motivational tendencies which are part of our animal nature. Consideration of specific cases is certainly not useless, but there is always a danger that our verdicts will be unduly influenced by feelings aroused by the details of the situation. Sidgwick expresses concern about the tendency to mistake subjective feelings for reasoned judgements. He puts the point thus: “It cannot be denied that any strong sentiment, however purely subjective, is apt to transform itself into the semblance of an intuition; and it requires careful contemplation to detect the illusion” (ME 339).³

The second essential feature of PHI is its foundationalism. Sidgwick believes that if an ethical principle is to be adequately justified, it must be supported by more fundamental propositions which are immediately cognizable in the abstract (ME 382).⁴ These fundamental propositions possess the attribute of self-evidence (the third feature of PHI) insofar as some degree of justification is obtained by reflecting on them alone. The apparently self-evident propositions are those that do not present themselves to us as needing further rational justification (EEM 29; ME 383).⁵ However, it is important to note that ‘self-evident’ does not mean ‘obvious’. Sometimes the self-evidence of a proposition can only be immediately

³ Sidgwick notes that the importance of abstraction for eliminating personal bias is the driving thought behind Kant’s Categorical Imperative as a test for rightness (ME 339, n. 1).

⁴ As the term is being used here, to say that a proposition is ‘cognizable’ is not to say that it is knowable, where this would imply something close to certainty. Rather, to cognize a proposition in the relevant sense is understand it and recognize a high likelihood of truth (ME 34).

⁵ One clear indicator that a principle is not self-evident is that it posits arbitrary definitions or boundaries (ME 293 n. 1).

recognized by those with certain background knowledge and conceptual competence. Although the suggestion that expertise is sometimes required for grasping self-evident truths might seem strange, Sidgwick points out that this is precisely how things work in the case of some of the axioms of mathematics and geometry. He puts the point thus:

Just as some mathematical axioms are not and cannot be known to the multitude, as their certainty cannot be seen except by minds carefully prepared,—but yet, when their terms are properly understood, the perception of their absolute truth is immediate and irresistible. Similarly, if we are not able to claim for a proposed moral axiom, in its precise form, an explicit and actual assent of “*orbi terrarium*,” it may still be a truth which men before vaguely apprehended, and which they will now unhesitatingly admit. (ME 229)

Sidgwick is well aware of the potential pitfalls of appeals to self-evidence in ethics.

First, some apparently self-evident propositions that seem to constitute substantive claims are really just concealed tautologies. Sidgwick gives an example of a purported principle of justice which states that “we ought to give every man his own.” This principle appears to have the abstractness, plausibility, and fundamentality of a self-evident axiom. However, upon reflection we see that it is uninformative because “his own” just means “that which it is right [that] he should have” (ME 375).⁶ A second danger is that the apparent self-evidence of a principle is sometimes merely the product of convention or repetition. A third is that many principles purported to be self-evident by a given theorist are inconsistent with the purportedly self-evident principles posited by other theorists.

In light of these concerns, Sidgwick posits four criteria that an apparently self-evident proposition must meet in order for it to possess the “highest certainty attainable.” This level of

⁶ An alternative reading of the principle implies that we must always return belongings to their owners. In “Further on the Criteria of Truth and Error,” Sidgwick points out that this is the principle that Polemarchus initially accepts, but comes to reject once he is convinced that it is not just to return a weapon to a friend who has fallen mad (EEM 166-7). Cf. Plato’s *Republic*, Bk. I. 331c-e.

justification must be approximately realized by our foundational ethical propositions if our reasoning is to lead to trustworthy conclusions (*ME* 338). The four criteria for highest certainty are as follows:

- (1) The terms of the proposition must be clear and precise.
- (2) The appearance of self-evidence must survive careful reflection.
- (3) The proposition must not conflict with other self-evident propositions.
- (4) The proposition must not be denied by an epistemic peer. (*ME* 338-41)

Admittedly, it is possible for a proposition that meets all four criteria to be false nonetheless. But the lack of absolute certainty is an unavoidable feature of our epistemic situation. Thus, Sidgwick insists that any proposition meeting these criteria possesses “as high a degree of certainty as [we] can hope to attain under the existing conditions of human thought” (*EEM* 109).

The primary motivation for the first criterion comes from the observation that the apparent certitude of a proposition can disappear once the key terms are rendered more precise. This point is illustrated in Sidgwick’s discussion of the common-sense precept that it is one’s duty to act wisely. Though this principle might strike us as self-evident, problems arise from the lack of precision. Read one way, the principle essentially tells us that we ought always to do what we have most reason to do. Though this is certainly true, it is no more useful than the claim that it is always right to do what is right. Read a second way, the principle tells us that we ought to cultivate a habit of always acting according to the verdicts of our reasoning, rather than allowing our actions to be determined by instincts or inclination. The problem here is that it is far from certain that this principle is true. Experience suggests that in some cases, rational ends are better attained by those who do not aim at them in a purely rational manner. To take one of Sidgwick’s examples, marriages seem more likely to succeed if they result from falling in love

rather than ‘tranquil and deliberate design’ (ME 345).⁷ Of course, Sidgwick believes that careful reasoning is indispensable for discovering our ultimate ends and fundamental ethical principles. He is simply making the point that the best means of achieving these ends and living in accordance with true principles may sometimes require following the influence of feelings and instincts rather than rational deliberation.⁸

Regarding the second criterion, the key insight is that reflection can help us recognize when the apparent self-evidence of a proposition is illusory. We sometimes judge principles to be self-evident simply because of their familiarity, or because they mesh with our inclinations or other non-rational impulses. In such cases we may be able to detect the error by considering whether our confidence in it might have originated from a pernicious source (e.g. racist principles believed as a result of being raised in a racist society). Crisp (2002: 71) helpfully summarizes this function of Sidgwick’s second test from a first-person perspective: “Ultimately I am seeking to vindicate my beliefs, to reassure myself that I believe that *p* because it is indeed the case that *p*.”

The third and fourth criteria explain why, despite the emphasis on foundational propositions, PHI is also partly coherentist. The degree of justification for a given belief depends in part on consistency with the agent’s other beliefs, as well as the beliefs of other equally reliable epistemic agents. Regarding the former, Sidgwick notes that ethicists have

⁷ Crisp (2006: 89) notes that Sidgwick’s first criterion does not merely test for clarity and precision in the meaning of the terms, the subject assessing the proposition must also be clear about the practical implications of the proposition. See Sidgwick (ME 215).

⁸ On this point Sidgwick adds the following: “Though the ‘dictates of Reason’ are always to be obeyed, it does not follow that ‘the dictation of Reason’—the predominance of consciously moral over non-moral motives—is to be promoted without limits...” (ME 395). Here we should read ‘moral’ and ‘non-moral’ as meaning ‘rational’ and ‘non-rational’. The suggestion that ultimate rational ends might be best promoted by allowing oneself to be influenced by non-rational impulses has been influential among contemporary consequentialists (see Driver: 2012).

frequently paid insufficient heed to the seriousness of internally inconsistent principles.⁹ He claims that such conflict is not a small problem to be put aside for future resolution, but is rather “absolute proof” that at least one of the principles must be qualified. Of course, there is nothing inherently problematic about qualified principles in ethics. But Sidgwick’s point is that adding qualifications requires reassessing the apparent self-evidence of the principle. Though the proposition might have had an air of self-evidence in its original form, the need for qualification, “suggests a doubt whether the correctly qualified proposition will present itself with the same self-evidence as the simple but inadequate one; and whether we have not mistaken for an ultimate and independent axiom one that is really derivative and subordinate” (ME 341).¹⁰ Here again reflection plays an important role. As Sidgwick notes in the first chapter of the *Methods*, many candidate first principles seem satisfactory to us when they have the field to themselves. But once we reflect and see potential conflicts with other plausible principles, the air of self-evidence diminishes (ME 14).

The fourth criterion anticipates contemporary concerns over the epistemic implications of disagreement. Sidgwick’s view is that a proposition cannot have the highest certainty attainable if an equally competent judge does not accept it. Such conflict does not warrant complete abandonment of the proposition, but it does require at least a temporary reduction in confidence. Sidgwick puts the point as follows:

⁹ Sidgwick does not clarify whether he is referring only to logical inconsistency, or also practical inconsistencies. Given his lengthy treatment of practical contradictions between egoism and utilitarianism, it is reasonable to suppose he has both types of inconsistency in mind. On this point see Crisp (2006: 90).

¹⁰ For further discussion and defense of the foundationalist-coherentist hybrid reading of Sidgwick, see Phillips (2011: ch.5).

For if I find any of my judgments, intuitive or inferential, in direct conflict with another mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality. And though the total result in any mind is not exactly suspense of judgment, but an alternation and conflict between positive affirmation by one act of thought and the neutrality that is the result of another, it is obviously something very different from scientific certitude. (ME 342)¹¹

It is important to note the qualifier in the above passage: “if I have no more reason to suspect error in the other mind than in my own.” When there are strong grounds for believing that the other party is mistaken, the existence of disagreement need not undermine one’s confidence in the proposition in question.

Having provided and defended his four criteria for highest certainty, Sidgwick is in position to provide his own candidate axioms and attempt to utilize them for the derivation of properly justified ethical principles.

3 Sidgwick’s Axioms of Ethics

3.1 Between Scylla and Charybdis

By the time we reach Sidgwick’s chapter on Philosophical Intuitionism in the *Methods*, he has already dedicated several chapters to the task of exposing the ways in which common-sense morality is epistemically suspect. The Dogmatic Intuitional system is plagued by inconsistencies, arbitrariness, and a lack of unifying explanations. And many of the common-sense principles that initially seem most certain turn out to be mere tautologies. Sidgwick memorably asks whether philosophical ethics might do better than this: “Can we then, between this Scylla and Charybdis of ethical inquiry, avoiding on the one hand doctrines that merely

¹¹ For contemporary discussion of the epistemology of disagreement see Kelly (2005), Audi (2010: ch. 14), and Christensen and Lackey (2013).

bring us back to common opinion with all its imperfections, and on the other hand doctrines that lead us round in a circle, find any way of obtaining self-evident moral principles of real significance?” (ME 379). Sidgwick answers this question in the affirmative, although his optimism is qualified. He believes there is a set of absolute, self-evident ethical principles, although they are too abstract to provide concrete guidance by themselves. Still, these principles at least have the potential to provide the stable foundation necessary for justification in ethics.

There is substantial controversy over the precise number and content of Sidgwick’s axioms. Part of the problem is that he is somewhat loose with terminology, occasionally oscillating between the labels ‘axiom’, ‘maxim’, ‘rational intuition’ and ‘principle’. A further problem is that some of the principles outlined in his discussion are said to be derived from more basic propositions, which suggests that they may not be genuinely self-evident.¹² As a result of these interpretive difficulties, a plausible case can be made for the number of genuine axioms being as small as three or as great as 11.¹³ As I interpret the relevant passages in the *Methods*, there are 5 distinct axioms from which three further principles are derived.

My preferred reading is motivated in part by Sidgwick’s remarks about what the various axioms and principles have in common. Of particular interest is his observation that the axioms and derived principles can be obtained by considering the relationship between parts and wholes (ME 380, 382-83). In some cases, the relevant relationship is the similarity of the individuals that make up a “Logical Whole or Genus.” By this Sidgwick means the entire group of

¹² For one discussion of this issue see Schneewind (1977: 294-5). The fact that Sidgwick derives some of his principles from other propositions does not rule out his viewing them as self-evident. Given his frequent references to the axioms of mathematics and geometry, Sidgwick clearly believes that proofs can be provided for self-evident propositions. It may be that Sidgwick thinks he can grasp the truth of his principles merely by reflecting on them, and the derivations are simply provided for those who might have greater difficulty cognizing them. For helpful discussion of self-evidence in ethics, and the point that self-evidence is compatible with provability, see Audi (2015).

¹³ For historical treatments of this debate see the references in Schneewind (1977: 290 n). For more recent treatments of, and contributions to, the debate over the axioms, see Skelton (2008), Irwin (2009 vol. 3, ch. 83); Phillips (2011: ch. 4), Shaver (2014), and Crisp (2015: 115-126).

individuals who possess the traits which make them such that certain normative concepts apply to them. For instance, all individuals who are capable of understanding and responding to reasons make up the logical whole of rational beings.¹⁴ The other type of whole that is relevant for the axioms is the “Mathematical or Quantitative Whole.” An example of a quantitative whole is a temporally instantiated and connected series of events such as the life of a particular person (Crisp 2015: 117, n. 37). According to Sidgwick, this idea is said to be indicated in the common notion of ‘the Good’ or ‘good on the whole for any individual human being (ME 381, my italics). Earlier in the *Methods*, Sidgwick characterizes ‘the good’ for a particular individual as that which she would have reason to desire on the assumption that only her own existence is to be considered (ME 112).¹⁵ This is contrasted with the notion of ‘good’ unqualified by reference to a particular subject. This latter notion is taken to refer to that which any rational being as such would have reason to desire on the assumption that *all* existence it to be given equal consideration. Sidgwick sometimes refers to this as ‘universal good’, and he suggests that this is the notion picked out when conduct is judged ‘good’ or ‘desirable’ in itself (ibid.).¹⁶

Sidgwick’s remarks about the relevance of parts and wholes for obtaining the axioms and derived principles allow us to see how they might be formulated and presented more

¹⁴ See Irwin 2009 vol. 3, ch. 83). Citing some of Sidgwick’s less well-known works, Crisp (2015: 117, n. 37) suggests that an example of a logical whole might be “the series of numbers, understood atemporally, each being a number as much as any other.”

¹⁵ In his discussion of an individual’s good in this passage, Sidgwick uses the phrase ‘ultimate good on the whole for me’ rather than simply ‘good for me’. The extra verbiage is motivated by recognition that something might be good for me if only a particular temporal part of my life were to be considered, yet not be good for me if my entire life were considered. This is an important point, but for ease of exposition I will use the phrase ‘the good for X’ to express the notion of X’s ultimate good on the whole.

¹⁶ Sidgwick occasionally uses the terms ‘absolute good’ and ‘objective good’ to express the idea of ‘universal’ or ‘intrinsic’ good. I have chosen to avoid using ‘absolute’ because this is sometimes taken to mean a type of good that is not good *for* anyone, such as unperceived beauty (see Kraut 2011). If Sidgwick’s ‘universal good’ exists, it might still be the case that nothing is good unless it is good *for* someone. I have chosen to avoid the term ‘objective good’ because this is sometimes taken to mean an end that a rational being ought to aim at independently of whether it would satisfy any of her desires. It is possible that such ‘objective good’ exists even though Sidgwick’s ‘universal good’ does not. This would be the case if, for instance, egoism is true. I will sometimes use ‘agent-neutral good’ as a substitute for ‘universal good’, because the former nicely captures the idea Sidgwick has in mind (though this terminology did not appear until Parfit’s *Reasons and Persons* (1984)).

systematically than they are in the *Methods*. My proposed formulations of the axioms and principles are as follows:¹⁷

Equal Right: The general requirements of ethics apply equally to all rational beings as such. (379-80)¹⁸

Rational Being: Qua member of the logical whole of rational beings, I am qualitatively identical to any other person. (379-80)¹⁹

From these two axioms we deduce the following practical principle:

Universal Right: If an action is right (or wrong) for me, it must also be right (or wrong) for any other rational being facing exactly similar circumstances. (379-80)²⁰

The axioms above are obtained by consideration of the logical whole of rational beings.

Sidgwick indicates that the other three are obtained by consideration of quantitative wholes rather than logical wholes:

¹⁷ None of my formulations match Sidgwick's wording exactly. The page numbers refer to the passages where the relevant axiom or principle is discussed. I defend my deviations from Sidgwick's wording below.

¹⁸ Sidgwick does not assign labels to the axioms, though he does give labels to the derived principles.

¹⁹ This axiom appears to be trivially true. This might seem problematic given Sidgwick's stated aim of discovering substantive principles that are unlike the tautologous "sham axioms" of common-sense morality. However, Sidgwick can achieve this aim as long as his derived principles are genuinely substantive and normative. And this requires only that at least one of the axioms from which it is deduced is substantive and normative. David Phillips (2011: 124) makes this point in defending his preferred reading of the axioms.

²⁰ Sidgwick notes that there is a corresponding principle stating that "it cannot be right for *A* to treat *B* in a manner in which it is wrong for *B* to treat *A*, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment" (ME 380). I do not consider this an independent principle because it is implied by the principle of Universal Right. If there is no relevant difference between *A* and *B*, then if it is right for *A* to treat *B* in a certain manner, it follows from Universal Right that it would have to be right for *B* to treat *A* in that manner should their roles be reversed.

Equal Good: One ought to have equal fundamental concern for equal portions of the good. (382)²¹

Temporal Irrelevance: Equal temporal parts of one's remaining life are exactly similar qua portions of one's own good. (381)²²

Personal Irrelevance: The equal good of any two individuals is exactly similar qua portions of universal good. (382)²³

From 'Equal Good' and 'Temporal Irrelevance' Sidgwick deduces a principle of prudence:

Prudence: One ought to have equal fundamental concern for equal temporal parts of one's remaining life. (381)

²¹ The 'ought' in the axioms should not be read as a moral 'ought' but rather as an 'ought' of ultimate practical reasons. As Crisp notes, though Sidgwick often utilizes distinctively moral concepts, his primary concern is to discover principles telling us what is ultimately reasonable, or what one has overall reason to do (Crisp 2015: 17). I have included the word 'fundamental' to express the point that the axioms are about ultimate reasons rather than derivative reasons. One's having equal fundamental concern for equal portions of the good is compatible with one's having derivative reasons to prioritize one portion over another.

²² Temporal Irrelevance does not appear to have the level of certainty Sidgwick thinks it does. In recent debates about the value of the shape of a life, some have argued that the later parts of one's life carry greater significance for one's overall wellbeing (Velleman 1991; Slote 1983; Temkin 2012). C.D. Broad (1930: 225) responds to this challenge for Sidgwick's axiom by noting that intuitions supporting the idea that a life's shape matters might be distorted by facts about the secondary pleasures and pains of memory and anticipation. For present purposes we can avoid delving into this debate because the key axioms for answering the Partiality Question are Equal Good and Personal Irrelevance.

²³ Sidgwick's formulation of the claim that nobody's good is more intrinsically valuable than anyone else's includes the following qualifier: "unless, that is, there are special grounds for believing that more good is likely to be realized in the one case than in the other" (ME 382). This qualifier is undoubtedly meant to capture the point that the quality and duration of the lives of two individuals would make a difference in the portion of universal good they constitute. But I believe the qualifier is also meant to accommodate the possibility that an individual's good might constitute a smaller portion of universal good if she has a bad character. Some evidence for this reading comes from Sidgwick's initial presentation of the impartialist principle in the opening chapter of the *Methods*. There he presents the impartialist as holding that "it cannot be reasonable to take as an ultimate and paramount end the happiness of any one individual rather than that of any other—at any rate *if equally deserving* and susceptible of it..." (ME 10, my italics). In my formulation Personal Irrelevance, the word 'equal' before 'good' is meant to accommodate both the point about quality and duration as well as the point about desert. The axiom leaves open the question of whether five minutes of Hitler's pleasure is of equal intrinsic value as five minutes of Gandhi's pleasure because their disparate characters might change the values.

From ‘Equal Good’ and ‘Personal Irrelevance’ Sidgwick deduces a principle of impartialism:

Rational Benevolence: One ought to have equal fundamental concern for the equal good of all individuals. (382)²⁴

3.2 *The Role of ‘Good’ in the Axioms*

An initial difficulty arises concerning the notion of ‘good’ and its place in the axioms. As we have seen, Sidgwick believes this term can be used to express two different ideas, depending on whether reference is made to a particular subject. If I say ‘the good *for x*’ where *x* represents a specific individual, I mean that which *x* ought to desire considering only *x*’s own existence.²⁵

We can call this the *personal specification* of ‘good’. If, on the other hand, I use this phrase without making reference to anyone in particular, I mean to refer to that which ought to be desired and promoted by all rational beings *as such*, regardless of how they are related to it.²⁶

We can call this the *universal specification* of ‘good’.

²⁴ One might worry that the requirement of equal fundamental concern expressed in Prudence and Rational Benevolence can be met by not having any concern for the respective ends. But since these principles are derived from axioms concerning ‘good’, and ‘good’ is understood as that which a rational being ought to aim at, Prudence and Rational Benevolence should be read as requiring positive concern for the respective ends.

²⁵ Crisp (2015: 61) points out that Sidgwick’s account of ‘good for’ is vulnerable to ‘the wrong kind of reasons objection’ (Rabinowicz and Ronnow-Rasmussen, 2004: 393). Crisp provides an example in which an evil demon threatens to torture you unless you desire something seemingly irrelevant to your good such as a saucer of mud. This threat would give you a strong reason to desire the saucer of mud, but it seems implausible that the mud is a constituent of your welfare. This is taken as a reason to reject any account, such as Sidgwick’s, that explains ‘good for’ in terms of what an agent ought to desire. Not much hangs on this for present purposes, as Crisp (2015: 61) suggests that Sidgwick could have avoided this difficulty by treating ‘good for’ as a primitive notion. Nonetheless, one might defend Sidgwick’s analysis by drawing on the distinction between ultimate reasons and derivative reasons. In the evil demon case, the reason to desire the saucer of mud is clearly derivative. Sidgwick’s account of ‘good for’ could seemingly avoid the objection if understood in terms of ultimate rather than derivative reasons.

²⁶ Sidgwick’s gloss on the idea of universal good is that of an end that all rational beings should desire and seek to realize assuming equal concern for *all* existence (ME 112). This way of expressing the idea is slightly misleading. It gives the impression that it is a conceptual truth that if universal good exists, everyone’s welfare constitutes an equal portion of it. But this is not a conceptual truth. To see why, recall that Sidgwick equates the idea of universal good with that of intrinsic value. There is nothing incoherent about believing that one’s own welfare is more intrinsically valuable than that of someone else (though this is certainly implausible). In order to avoid the implication that everyone’s welfare *necessarily* constitutes equal portions of universal good, I characterize universal good as that which ought to be desired by all rational beings as such, regardless of how they are related to it. I take it that this captures what Sidgwick has in mind. Some evidence for this reading is found in *Methods* IV: II where

As I have presented them, the three relevant axioms differ regarding ‘good’. Whereas the axioms of Temporal Irrelevance and Personal Irrelevance include personal and universal specifications respectively, the axiom of Equal Good includes an underspecified usage. It may seem puzzling that ‘good’ could be used without full specification, given Sidgwick’s suggestion that a usage without reference to any particular subject means “what as a rational being I should desire and seek to realize, assuming myself to have equal concern for *all* existence” (ME 112). However, Sidgwick’s understanding of ‘good’ and the two different conceptions actually allows for a usage that maintains neutrality between the personal and universal specifications. On this alternative usage, ‘good’ means ‘that which any rational being would have reason to desire and seek to realize’. It is left open whether this end necessarily includes only one’s own welfare (because ‘good’ is agent-relative), or if it might extend beyond oneself (because ‘good’ is universal).²⁷

I believe it is this neutral usage of ‘good’ which occurs in the axiom of Equal Good. Before considering the points in favor of this reading, note first that further specification is not needed for the axiom to have *prima facie* plausibility. This can be seen by substituting the neutral gloss on ‘good’ in my suggested formulation: “One ought to have equal fundamental concern for equal portions of that end which, as a rational being, one ought to desire and seek to realize.” This is a substantive proposition, and contemplation of it alone provides at least preliminary justification for its acceptance.

Further motivation for my underspecified reading of Equal Good comes from the following considerations. First, notice that if the axiom were intended to be read with an agent-

Sidgwick equates the idea of absolute, unqualified good with the idea of an end “to which the action of a reasonable agent as such ought to be directed” (ME 421). The ‘as such’ in this statement is meant to encompass the idea that facts about how the agent happens to be related to the end in question are normatively inert.

²⁷ One might worry that the neutral usage would render the axiom of Equal Good incapable of passing Sidgwick’s first test which requires clarity and precision. I shall address this point presently.

relative specification of ‘good’, Sidgwick could not have derived the impartialist principle of Rational Benevolence from Equal Good and Personal Irrelevance. Impartialism does not follow from (1) the fact that one ought to have equal fundamental concern for equal portions of the good, and (2) the fact that everyone’s good is of equal intrinsic value. Hence, a reading of Equal Good with an agent-relative specification of ‘good’ is ruled out. Thus, the only live alternative to the underspecified reading I favor is a reading according to which ‘good’ takes a universal specification. One reason for favoring the underspecified reading over the universal reading is that the underspecification of ‘good’ allows for Equal Good to be accepted by egoists and non-egoists alike. This would not be the case if the axiom contained a universal specification of ‘good’. If the axiom contained such a specification, it would imply that the ultimate end of rational conduct is not relative to each individual. The egoist adamantly denies this. When considering candidate versions of genuinely self-evident axioms it is preferable to choose a version that is accepted by all parties to the debate because substantial disagreement diminishes the degree of certainty.²⁸

This is related to a further point in favor of the underspecified reading which is that Sidgwick himself has substantial doubts about whether the ultimate end of rational conduct is universal rather than agent-relative. These doubts about the existence of a non-relative end for all rational beings as such are precisely what lead to his dissatisfaction with Mill’s defense of utilitarianism. And as I shall explain presently, these same doubts also lead to the famous conclusion that there is a Dualism of Practical Reason. Hence, it is doubtful that Sidgwick would find it self-evident that rational agents are bound to aim at *universal* good.

²⁸ Robert Shaver (2014) attempts to explain how egoists can accept Equal Good by interpreting it as having an unstated antecedent of ‘from the point of view of the universe’. On this reading, the egoist can accept the axiom as a truth about what would follow were one to adopt a universal point of view. The egoist can still avoid being driven to impartialism by simply refusing to adopt such a point of view. One advantage of my reading is that it allows for the egoist’s assent without having to posit an unstated antecedent.

An additional reason for reading Equal Good with an underspecified usage of ‘good’ is that this fits well with Sidgwick’s remarks about the usefulness of his axiom-based argument against proponents of egoism. As will become clear in my discussion of the Dualism of Practical Reason, Sidgwick does not believe his derivation of Rational Benevolence has dialectical force against all proponents of egoism. Specifically, he does not believe his argument is of any use against those egoists who believe all good is relative to the individual. However, Sidgwick does believe his axioms are useful against an egoist who acknowledges the existence of universal good. The suggestion is that if an egoist admits that his own welfare is part of universal good, it then becomes pertinent to note that his own good cannot be a greater part of universal good than anyone else’s. From here Sidgwick claims that, “starting with his own principle,” the egoist may be brought to accept that the happiness of all is an end that the action of rational beings as such ought to be directed (ME 421). One plausible way of understanding Sidgwick here is that all egoists agree that one ought to have equal fundamental concern for equal portions of the good. They can resist impartialism as long as they deny that the good is universal. But if an egoist acknowledges universal good, his acceptance of Equal Good combined with the truth of Personal Irrelevance will lead him directly to impartialism.

Though these considerations tell in favor of the underspecified usage of ‘good’, two further worries remain. The first is that the lack of specification in Equal Good seems to tell against its status as an axiom of the highest certainty attainable. As we have seen, Sidgwick’s first criterion for highest certainty requires clarity and precision. The underspecification of ‘good’ in my formulation seems to make it a prime example of imprecision.

In order to properly assess this worry it is necessary to recall the motivation for the ‘clear and precise’ criterion. As we have seen, Sidgwick pays considerable heed to the fact that many

widely accepted ethical principles only appear self-evident as long as their terms remain vague. Recall the earlier example of a principle stating that it is one's duty to act wisely. We saw that this principle becomes highly dubious upon clarification. Understood one way, the principle tells us only that it is our duty to do what we judge to be our duty. Understood another way, the principle tells us that we ought to cultivate a habit of acting strictly from the deliverance of reason rather than allowing our actions to be determined by instincts and emotions. While the first precisification renders the principle devoid of substance, the second substantially weakens its plausibility. This type of problem illustrates the purpose of the first criterion. Aiming for clarity and precision allows us to uncover serious problems afflicting principles that might appear self-evident at first blush.

The important point for present purposes is that the underspecification of Equal Good is not the sort of problematic imprecision that the first criterion tests for. To see why, notice that neither of the candidate specifications of the principle renders it dubious. Equal Good has the interesting feature of being equally plausible no matter which way of specifying it corresponds to the way the world is. Whether the good is agent-relative or universal, it appears evident that I ought to have equal fundamental concern for equal portions of it. When I reflect on the notion of an end that I ought to desire and seek to promote, it seems clear that I ought to give equal weight to equal portions of it. I can be confident about this even though I might be uncertain about whether this quantitative whole is agent-relative or agent-neutral.²⁹

That being said, one might still worry that even if the lack of specification does not render Equal Good dubious, it does diminish its usefulness. More specifically, one might have doubts about whether we can make valid inferences from the axioms to the principles of

²⁹ Unfortunately, the axiom of Equal Good does run afoul of Sidgwick's fourth test. As I shall discuss below, some of Sidgwick's most loyal contemporary disciples implicitly reject the axiom.

Prudence and Rational Benevolence if the axioms differ in their specifications of the key term. Fortunately, the divergence in meaning that occurs in the axioms is not the type of ambiguity that leads to the fallacy of equivocation. When a key term is underspecified in one premise and specified a particular way in another, a valid inference can still be made if the conclusion would follow on either specification of the term in the original premise. This is exactly what happens in the case of the deduction of Prudence from Equal Good and Temporal Irrelevance. We have seen that while Equal Good includes an underspecified usage of ‘good’, Temporal Irrelevance includes a personal specification. The reason we can still deduce Prudence from these axioms is that the personal specification in Temporal Irrelevance will have the same significance no matter which account of ‘good’ turns out to be veridical. The account that is veridical is the one that corresponds to the way the world actually is. If the world is such that what is ultimately desirable for rational beings is always relative to the individual, then the personal account is veridical. If the world is such that what is ultimately desirable for rational beings is universal rather than agent-relative, then the universal account is veridical. Though Equal Good is given a neutral specification, the referent of this axiom is fixed by the facts about the world. The axiom refers to whichever account of ‘good’ is veridical.

Suppose it turns out that good is always relative to the individual. If so, then Equal Good would refer to personal good just as Temporal Irrelevance does. Hence the deduction of Prudence would not be problematic at all. But suppose instead that the world is such that the ultimate end for any rational being is universal rather than personal. In that case, Equal Good would refer to universal good. This might seem like a problem given that Temporal Irrelevance is about personal good. However, on the reasonable assumption that the personal good of the individual would constitute at least a portion of universal good, it would still be the case that the

individual ought to promote her own good with equal regard for equal portions of it. Hence, when this requirement is combined with the fact that the different temporal parts of a person's life are the same qua portions of her good (i.e. the axiom of Temporal Irrelevance), we can conclude that the individual ought to be impartial with respect to the different temporal parts of her remaining life (i.e. the principle of Prudence).

Alas, things do not work quite so nicely for the deduction of Rational Benevolence from Equal Good and Personal Irrelevance. Rational Benevolence only follows from these axioms if it turns out that the ultimate end for rational beings is universal rather than personal. If 'good' is universal, then Equal Good refers to the same notion that Personal Irrelevance picks out. Hence, there would be no trouble deriving the conclusion that reason requires impartiality at the fundamental level. However, if 'good' is always relative to the individual, then Equal Good refers to a notion that is different from that picked out by Personal Irrelevance. Thus, Rational Benevolence could not be derived from the axioms. This may seem to undermine any potential usefulness of Personal Irrelevance for answering the Partiality Question. However, the axiom is only useless if the idea of universal good cannot be vindicated. If it can be vindicated (and I shall argue that it can), then we can use Equal Good and Personal Irrelevance to derive a principle of impartialism after all. As we shall see presently, Sidgwick is well aware of the fact that the usefulness of his axioms hinges on the question of whether the good is universal or agent-relative.

3.3 Sidgwick's Words

I take the foregoing considerations to constitute an adequate defense of my usage of 'good' in formulating the axioms. But given the importance of Equal Good and Personal Irrelevance for

Sidgwick's project, as well as for my argument for Rational Impartialism, it is necessary to examine Sidgwick's own statements of these axioms in order to fully vindicate my formulations. Since I have opted to deviate from Sidgwick's wording, I take on the burden of demonstrating that I have captured his intended meaning nonetheless.³⁰

After explaining how the principle of Prudence is obtained by considering the similarity of the different temporal parts of the individual's good on the whole, Sidgwick points out that similar reasoning can be applied in the case of universal good in order to derive the principle of Rational Benevolence. The key passage reads as follows:

And here again, just as in the former case [i.e. the case of an individual's good], by considering the relation of the integrant parts to the whole and to each other, I obtain the self-evident principle that the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other; unless that is, there are special grounds for believing that more good is likely to be realized in the one case than in the other. And it is evident to me that as a rational being I am bound to aim at good generally,—so far as it is attainable by my efforts,—not merely at a particular part of it. (ME 382)

Sidgwick adds that from these two axioms we can deduce the principle of Rational Benevolence, according to which “each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him” (ME 382).

The statement from which I draw the axiom of Personal Irrelevance is, “the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other...” The obvious difference between this statement of the axiom and my preferred formulation is that whereas my formulation emphasizes similarity qua

³⁰ Of course, I am not alone in taking on this burden. The need to deviate from Sidgwick's wording in order to fully clarify his intended meaning is widely accepted among Sidgwick scholars.

portions of universal good, Sidgwick’s statement mentions equal importance from the point of view of the universe. Despite this surface difference, I believe my formulation captures exactly what Sidgwick has in mind. My primary reason for emphasizing similarity qua portions rather than equal importance is that this better accommodates Sidgwick’s claim that the same process of reasoning leads to the three derived principles—Prudence, Rational Benevolence, and Universal Right.³¹ The process of reasoning that leads to Universal Right and Prudence is clearly based on the similarity of different parts qua portions of the respective logical and quantitative wholes. By formulating Personal Irrelevance as I have, the structural similarity between these deductions and the deduction of Rational Benevolence is rendered transparent. Moreover, the claim that the equal good of any two individuals is exactly similar qua portions of universal good is equivalent to the claim that neither person’s good is of greater importance from ‘the point of view of the universe’. The idea of something being important from the point of view of the universe is meant to capture the notion of an end that is desirable for all rational beings *as such*, regardless of how they are related to it. This is the same notion picked out by ‘good’ with a universal specification.³²

My formulation of Equal Good is more controversial. Sidgwick’s statement of this axiom is located in the part of the above passage that reads “as a rational being I am bound to

³¹ Sidgwick refers to the principle I call ‘Universal Right’ as the principle of ‘Justice’ (ME 382).

³² Crisp suggests that the point of view of the universe is simply the perspective one takes when considering all sentient beings as a group (2015: 122, n. 51). While it is true that the quantitative whole of universal good includes the welfare of all, textual evidence suggests that the notion of something being good from the point of view of the universe involves more than it merely being part of the sum total of welfare in the world. In his essay, “Mr. Barratt on ‘The Suppression of Egoism,’” Sidgwick claims that when something is ‘desirable from a universal point of view’, this is equivalent to its being ‘absolutely’ good, or ‘what all rational beings, as such, ought to aim at realizing’ (EEM 27). As I explain in footnote 17 above, for Sidgwick, ‘absolute good’ is equivalent to ‘universal good’ or ‘intrinsic good’. Note that the axiom of Personal Irrelevance is a substantive claim. It is conceptually possible that one person’s happiness is more intrinsically (or absolutely) valuable than another’s. If this were the case, then all rational beings as such would have ultimate reason to give priority to that person’s happiness to the extent that it has greater intrinsic value. The axiom of Personal Irrelevance is the substantive claim that this possibility does not obtain because nobody’s happiness (or deserved happiness) is of greater intrinsic value than the equal happiness (or deserved happiness) of anyone else.

aim at good generally... not merely at a particular part of it.” It is not immediately obvious that this should be taken to mean “One ought to have equal fundamental concern for equal portions of the good.” As is the case with Personal Irrelevance, one obvious difference between my formulation of Equal Good and Sidgwick’s statement of it is my emphasis on equal portions. As before, the motivation for this emphasis is that it makes the similarity of structure among the deduction of all the intermediate principles more readily apparent. All of the deductions are based on abstract maxims concerning quantitative or logical wholes, and the similarities among different parts of those wholes.

That being said, it is certainly tempting to read Sidgwick’s statement of Equal Good as an assertion that one ought to aim at the good of everyone rather than just one’s own good.³³ This reading seems especially compelling if we take ‘bound to aim at good generally’ to be synonymous with ‘bound to aim at the general good’. But we have already encountered two reasons for thinking that this small difference in word order marks an important difference in meaning, and that it is no accident that Sidgwick says ‘good generally’ rather than ‘general good’. First, given some of Sidgwick’s commitments, it would be strange for him to classify a principle stating that one ought to aim at general good (i.e. the good of everyone) among the self-evident axioms. He certainly considers egoism to be a viable option, and at least some proponents of egoism are thereby epistemic peers. Hence, the egoist’s insistence on the falsity of

³³David Phillips (2011: 124) favors a reading of the above passage that renders Equal Good less significant. On Phillips’ reading, the axiom is merely an expression of the idea that there is reason to aim at whatever is good. Phillips claims that this makes Equal Good tautologous, and that partly for this reason, the deduction of Rational Benevolence cannot succeed even on the assumption that ‘the good’ is universal. The worry is that even if there is *some* reason to aim at whatever is universally good, reason might still permit prioritizing some parts of universal good over others. I take it to be an advantage of my reading of Equal Good that this problem is avoided. As I read Sidgwick, his claim is not just that there is some reason to aim at whatever is good but rather that one ought to have equal fundamental concern for equal portions of the good. Phillips would likely respond that on my reading the axiom is not genuinely self-evident because it is not obviously contrary to reason to prioritize one’s own good over the good of others, even if the good is universal rather than agent-relative. I shall return to this issue in chapter 4 where I consider a particularly incisive version of this challenge arising from the views of Roger Crisp (2006) and Derek Parfit (2011).

an axiom stating that one ought to aim at the good of all would diminish the appearance of self-evidence.³⁴ Second, if Sidgwick finds it genuinely self-evident that one ought to aim at the good of all he would not reach the troubling conclusion that egoism is as well justified as impartialism.

A further problem with the alternative reading of Equal Good mentioned above is that it cannot be combined with Personal Irrelevance to yield Rational Benevolence. Personal Irrelevance essentially states that nobody's welfare is more intrinsically valuable than anyone else's. It does not follow from this and the (ostensible) fact that one ought to aim at the good of everyone, that reason requires giving *equal* weight to the equal good of all. One might consistently accept the two axioms (on the suggested readings) while also believing that one ought to prioritize the good of some over the equal good of others. Proponents of the 'aim at everyone's good' reading of Equal Good might respond by proposing that the axiom be read as stating a rational requirement to aim at everyone's good equally. But the problem here is that this is essentially the principle of Rational Benevolence, which is the conclusion that Sidgwick aims to derive from Personal Irrelevance and Equal Good.

Before moving on to discuss Sidgwick's considered views about the epistemic status of egoism and Rational Benevolence, there is a final potential problem for my reading of the axioms that bears mentioning. It might appear as though one of the primary objections I raise against alternative readings of Equal Good also cuts against my preferred formulation of Personal Irrelevance. Just as the egoist would reject Equal Good if it implied that rational concern is not relative to oneself, so too should she reject Personal Irrelevance. This is because

³⁴ One might object that the same could be said for Sidgwick's principle of Rational Benevolence. At times Sidgwick suggests that this principle is self-evident, yet clearly it is rejected by the egoist. The solution to this puzzle is that Sidgwick only considers Rational Benevolence to possess *apparent* self-evidence. Immediately following his deduction of the intermediate principles of Prudence and Rational Benevolence, Sidgwick claims only to have shown that they contain "at least a self-evident element" (ME 382). The *genuinely* self-evident elements in Rational Benevolence are the axioms of Personal Irrelevance and Equal Good. On the reading I am advocating neither of these axioms conflict with other apparently self-evident propositions.

proponents of egoism deny the existence of the non-relative, 'universal' good implied by Personal Irrelevance. However, though the egoist claims that all good is agent-relative, she does not insist on the falsity of the claim that the good of any two individuals must be exactly similar qua portions of universal good. The egoist can still accept the latter as a truth we can grasp by reflecting on the concept of universal good and the relation this notion might have to each individual's good.³⁵ For it seems difficult to imagine grounds for believing that one's welfare could be any greater or lesser a part of universal good than the welfare of others would be, even if one has doubts about whether universal good exists at all. And importantly, Personal Irrelevance does not have ethical implications such that the egoist is committed to their falsity. Things are different in the case of the alternative reading of Equal Good we have been considering. On this alternative, the axiom implies that the ultimate aim of rational conduct is not relative to the individual. The egoist rejects this implication as false, insisting that rational concern is exclusively self-regarding. It is this sort of substantive ethical disagreement that precludes highest certainty. Of course, the fact that Personal Irrelevance makes reference to a type of good for which there is controversy regarding its existence is cause for concern. But we can at least be confident that *if* universal good exists, no individual's good constitutes a greater or lesser portion of it.

These considerations also explain why Sidgwick himself finds Personal Irrelevance genuinely self-evident despite his being tempted by egoism. He is certain that one ought to have equal fundamental concern for equal portions of the good (no matter which specification of

³⁵ On this point I am generally in agreement with C.D. Broad (1930: 244) and Robert Shaver (2014: 176). Note that on my reading, Personal Irrelevance is not tautologous. The axiom essentially states that nobody's welfare is more intrinsically valuable than anyone else's. Notice that if someone were to claim that her own welfare is more intrinsically valuable than that of others, she would not have said anything contradictory or incoherent. Nor would she betray conceptual confusion. We would reject her claim not because it is incoherent, but rather because it is deeply implausible and it wears its arbitrariness on its sleeve.

‘good’ corresponds to the way the world is). And he is confident that nobody’s good could constitute a greater or lesser portion of universal good than anyone else’s. What Sidgwick is uncertain about is whether ‘good’ really is universal, or if it is agent-relative as egoists maintain. It is this uncertainty which leads to the worrisome Dualism of Practical Reason. Before turning to my axiom-based argument for Rational Impartialism, it will be instructive to take a closer look at Sidgwick’s most notorious thesis.

4 Chaos in the Cosmos of Duty: The Dualism of Practical Reason

4.1 ‘Foredoomed to Inevitable Failure’

Following his derivation of Rational Benevolence from two of his axioms, Sidgwick claims that this principle is “required as a rational basis for the Utilitarian system” (ME 387). Rational Benevolence is only a basis for utilitarianism rather than a complete vindication of it because the principle does not tell us to maximize hedonic welfare in the aggregate by any means necessary. The principle tells us only that we have equal ultimate reasons to promote the equal good of all. This is in principle compatible with the existence of deontological constraints on the promotion of impartial good.³⁶ However, the fact that a principle requiring impartiality between self and others is derived from self-evident axioms should presumably rule out egoism as a viable option.

³⁶ Thomas Hurka (2014) argues that Sidgwick was guilty of equivocation in his treatment of the principle of Rational Benevolence. Hurka claims that Sidgwick must use the principle in a weaker ‘other-things-equal’ form in order to meet the consensus criterion, but that he must also use it in a stronger ‘all-things-considered’ form in order for it to support consequentialism and generate the Dualism of Practical Reason. As I read him, Sidgwick’s principle of Rational Benevolence falls somewhere in between Hurka’s two alternatives. It clearly is not the stronger version because it does not state that impartial welfare is the only source of one’s reasons. But my reading is stronger than Hurka’s weaker version because the ‘other-things-equal’ qualifier is too permissive—it might allow one to appeal to relational facts as grounds for having different fundamental concern for equal portions of the good. On my interpretation, the deduction of Rational Benevolence is intended to settle only the question of whether we have ultimate reasons of partiality. What the principle leaves open is whether there might be other fundamental principles (e.g. deontological constraints) that place limits on the pursuit of impartial welfare. My preferred ‘intermediate’ reading of Rational Benevolence is sufficient for providing an intuitive basis for consequentialism because it informs us that we ought to promote the good impartially. If, as Sidgwick initially believes, this principle

But to his great disappointment, Sidgwick eventually concludes that the axioms cannot even do this much. While there are no problems with Equal Good and Temporal Irrelevance, difficulty arises when we reach the axiom of Personal Irrelevance. It is hard to see grounds for denying that *if* that which rational beings *as such* ought to desire is agent-neutral rather than agent-relative, then the equal good of any two individuals would constitute equal portions of this end. Hence, if we could be confident about the universality of good, then Personal Irrelevance would be a useful principle. However, it is far from clear that the good is agent-neutral. And if the good is always relative to a particular individual, then Personal Irrelevance has no practical significance. The fact that nobody's good could constitute a greater portion of universal good becomes irrelevant if no such good exists.

Sidgwick is keenly aware of this problem, and this is why he claims that his argument for Rational Benevolence is of no use against an egoist who begins with the assumption that rationality requires pursuit of one's own good exclusively. He puts the point thus:

If the Egoist strictly confines himself to stating his conviction that he ought to take his own happiness or pleasure as his ultimate end, there seems no opening for any line of reasoning to lead him to Universalistic Hedonism as a first principle; it cannot be proved that the difference between his own happiness and another's happiness is not *for him* all-important. (ME 420)

In the final chapter of the *Methods*, Sidgwick further articulates his concern about the inability to undermine the egoist's position:

is the only one to survive critical scrutiny, the result is consequentialism. I disagree with Hurka's claim that only an 'all-things-considered' rendition of Rational Benevolence can give rise to the Dualism of Practical Reason. I explain how Sidgwick arrives at his dualism in the main text of this section.

It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently “I” am concerned with the quality of my existence as in individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual. (ME 498).³⁷

Because he finds egoism independently plausible, and because he sees no way of vindicating belief in the existence of universal good, Sidgwick concludes that egoism is as well justified as impartialism.

Though the equal standing of apparently contradictory principles suggests that a rational basis for ethics is unattainable, Sidgwick considers whether this conclusion might still be avoided. This would be possible if it could be demonstrated on empirical grounds that, as many moralists and theologians have maintained, the ends of self-interest and the greater good harmonize such that they never conflict with one another. Sidgwick explores this possibility and acknowledges that in many respects the two candidate ends are compatible. Given the facts of human nature, having a high level of regard for the welfare of others tends to contribute greatly to one’s own happiness. Those who are predominantly selfish miss out on the “sense of elevation and enlargement given by wide interests,” as well as the “more secure and serene satisfaction that attends continually on activities directed towards ends more stable in prospect than an individual’s happiness can be” (ME 501).

³⁷ David Phillips (2011: 127) suggests that, despite the wording of this passage, Sidgwick should be read as making a normative claim that one *ought* to have greater concern for the quality of one’s own existence, rather than the descriptive claim that the one *is* more concerned about the quality of one’s own existence. Phillips believes this to be the case because if Sidgwick were merely making a descriptive claim, this would not lend any support to egoism. However, as I read him, Sidgwick’s point is not that common-sense suggests that one ought to have greater fundamental regard for oneself. What common sense affirms is the real metaphysical distinction between persons, and the fact that most people do in fact have greater regard for themselves. Sidgwick’s worry is that it seems reasonable for the egoist to insist that the real distinction between persons provides grounds for prioritizing one’s own good.

Unfortunately, the overlap between the ends of one's own happiness and universal happiness is far from perfect. There are many circumstances in which promoting the general good requires a sacrifice of one's own good. Given these practical conflicts, and given the equal justification for principles of Egoism and Rational Benevolence, Sidgwick reaches the somber conclusion that there is an irresolvable "Dualism of the Practical Reason."³⁸ The upshot is that in a wide range of circumstances practical reason cannot serve as a useful guide. Sidgwick laments that in cases of perceived conflict between self-interest and the greater good: "practical reason, being divided against itself, would cease to be a motive on either side; the conflict would have to be decided by one or other of two groups of non-rational impulses" (ME 508). He puts the point even more dramatically at the conclusion of the first edition of *The Methods*: "The Cosmos of Duty is thus really reduced to a Chaos: and the prolonged effort of the human intellect to frame a perfect ideal of rational conduct is seen to have been foredoomed to inevitable failure" (ME1 473).

4.2 Avoiding Cosmic Chaos

Though I accepted Sidgwick's bleak conclusion upon my first exposure to the underlying ideas, I have since regained hope. My primary aim in this section is to explain how two of Sidgwick's axioms will be used as part of my argument for Rational Impartialism. If this argument succeeds, the failure of prolonged efforts in philosophical ethics will be prevented (or at least forestalled).

³⁸ In "Some Fundamental Ethical Controversies," Sidgwick expresses dissatisfaction with the term 'Dualism of the Practical Reason': "I am not particularly pleased with the phrase, which has a pretentious sound, and is perhaps liable to mislead by suggesting that I claim for my view a completeness of systematic construction which, on the contrary, I wish to avoid claiming; but it seemed the most convenient phrase to express the conclusion in which I was forced to acquiesce after a prolonged effort to effect a complete systematization of our common ethical thought" (EEM 43).

The two key axioms are Equal Good and Personal Irrelevance. We have seen that the Dualism arises from uncertainty about universal good. If we lack justification for believing that such good exists, then we cannot deduce a principle of impartialism from Equal Good and Personal Irrelevance. Hence, in order to vindicate Rational Impartialism, I shall argue that we are justified in believing that the good is not relative to individuals but rather comprises the good of all. This claim will constitute one of the premises in my argument for RI.

But even if the idea of universal good is vindicated, this will still not be enough to deduce Rational Impartialism. This is because it is *prima facie* plausible that partiality itself (e.g. loving relationships and the acts they give rise to) might be part of universal good. If partiality is valuable for its own sake, then we can accept Personal Irrelevance and a universally specified axiom of Equal Good while still maintaining that we have fundamental reasons to prioritize our loved ones. This would be the case, for instance, if the value instantiated in acts of partiality towards one's intimates could outweigh the value realized by the promotion of a stranger's well-being. Sidgwick does not worry about this because he takes himself to have provided sufficient justification for a purely hedonistic value theory (hence ruling out the intrinsic value of partiality).³⁹ Since the truth of hedonism remains highly controversial, I shall attempt to vindicate impartialism via a (somewhat) less contentious route. I shall argue that partiality is only derivatively good—that its value derives from the contributions it makes to welfare. This will constitute the final premise in my argument for Rational Impartialism. With the added premises, the argument for Rational Impartialism is as follows:

³⁹ Sidgwick's case for hedonism in *The Methods* is located in Book III, ch. XIV. See also his "Hedonism and Ultimate Good" (EEM 89-98).

1. One ought to have equal fundamental concern for equal portions of the good.
[Axiom of Equal Good]
2. The good is not relative to individuals but rather is universal.
3. One ought to have equal fundamental concern for equal portions of universal good.
(from 1-2)⁴⁰
4. The equal good of any two individuals is exactly similar qua portions of universal good. [Axiom of Personal Irrelevance]
5. Partiality towards one's intimates is only derivatively good.
6. One ought to have equal fundamental concern for the equal good of all individuals.
(from 3-5)⁴¹

If adequate support for premise 2 can be given, this will settle the question of how to specify 'the good' in the axiom of Equal Good. Whereas its initial presentation requires an underspecified reading, if we are justified in believing that the good is universal, we can be justified in accepting a universally specified version of Equal Good (premise 3). I shall provide arguments in support of premise 2 in the next chapter (chapter 3). In that chapter I shall also provide a defense of premise 5—the claim that instantiations of partiality are only derivatively good.

Given what has been said so far, one might think that my case for Rational Impartialism depends solely the strength of the arguments I will present for universal good and the derivative value of partiality. It certainly seems that this would have appeared sufficient to Sidgwick. This

⁴⁰ One might worry that a universal specification renders Equal Good tautologous. This worry seems especially compelling in light of Sidgwick's gloss of 'universal good' as that which rational beings ought to desire and seek to realize assuming equal concern for all existence (ME 112). However, one could coherently maintain that (1) there is an end such that all rational beings as such ought to desire and seek to realize it, (2) this end is that which a rational being ought to desire assuming equal concern for all existence, and (3) reason does not require equal concern for equal portions of this end. That this is not an inconsistent triad can more readily be seen when we remember that 'universal good' is synonymous with 'intrinsic good'. It is at least coherent to claim that one ought to prioritize a smaller portion of intrinsic value over a larger portion. I shall return to this issue as part of my defense of the axiom of Equal Good in chapter 4.

⁴¹ Just as in the axiom of Equal Good, the 'ought' in the conclusion of this argument should not be read as a moral 'ought' but rather as an 'ought' of ultimate practical reasons.

is because the other two premises are constituted by genuine axioms, and the difficulty concerning Equal Good would be eliminated with the vindication of premise 2. However, one problem that Sidgwick could not have anticipated is that two of his most prominent disciples reject the axiom of Equal Good. Derek Parfit (2011) and Roger Crisp (2006) independently champion a view which I call the Partial Dualism of Practical Reason. On this view, although everyone's welfare constitutes an equal portion of universal good, agents sometimes have ultimate reason to prefer their own lesser good to the greater good of others.⁴² The explanation for the extra weight of self-interest is that the intrinsic value of an individual's welfare is not the only factor that determines our reasons to promote it—it also matters how we are related to that person. Both Parfit and Crisp maintain that the relations we have to our own consciousness justify giving greater weight to our own welfare than the welfare of others, even though everyone's welfare constitutes an equal portion of the good. I shall thus dedicate chapter 4 to defending Equal Good against the challenges presented by Parfit and Crisp.

5 Conclusion

In this chapter I have outlined the details of Sidgwick's 'Philosophical Intuitionism', and explained why this is an attractive approach to ethical inquiry. I have also provided novel formulations of Sidgwick's axioms, and defended these formulations against various objections. Lastly, I explained how Sidgwick reaches the conclusion that there is an irresolvable Dualism of Practical Reason, and I outlined my strategy for avoiding this bleak view through a Sidgwick-inspired argument for Rational Impartialism. In the next chapter I provide arguments for two highly controversial claims about the good—that it is agent-neutral, and that partiality is only

⁴² An early articulation of Partial Dualism is found in Broad's chapter on Sidgwick in *Five Types of Ethical Theory* (1930: 243-6).

derivatively part of it. These two claims respectively constitute the second and fourth premises of my argument for RI.

3 Universal Good

1 Introduction

As we saw in the previous chapter, Sidgwick's Dualism arises from uncertainty about the scope of 'good'. The axiom of Equal Good states that one ought to have equal fundamental concern for equal portions of the good. Sidgwick's problem was that he could not see any way of settling the question of whether anything is good 'from the point of view of the Universe', or if instead there is only agent-relative good. If the good is always relative to the individual, as egoists maintain, then Sidgwick's axioms support egoism rather than impartialism. One of my aims in this chapter is to make a compelling case for believing that universal good does in fact exist. If we can be justified in believing that the good is universal rather than agent-relative, then the axiom of Equal Good can be read with a universal specification. This would in turn allow for the possibility of deriving a principle of impartialism from Sidgwick's axioms after all. But as I mentioned previously, securing the impartialist conclusion requires arguing against the purported intrinsic value of partiality towards loved ones. Hence, a second aim of this chapter is to argue that instantiations of partiality are not good in themselves, but are rather of mere derivative value.

The discussion proceeds as follows. I begin in section 2 by further explicating the idea of universal good, and arguing that the best test for the existence of universal good involves consideration of cases in which promoting the interest of others requires no sacrifice on the part of the agent in question. I present such a case, and I explain why it lends substantial support to the existence of universal good. I also explain why the denial of universal good looks especially implausible on the assumption that practical reasons are object-given and value-based. In section

3 I turn to the question of whether instantiations of partiality within the context of loving relationships are intrinsically valuable. I argue that partiality's value is merely derivative from the contributions it makes to welfare. A brief conclusion follows.

2 The Point of View of the Universe

2.1 The Idea of Universal Good

Recall that Sidgwick characterizes the ultimate good for the individual as “what I should practically desire if my desires were in harmony with reason, assuming my own existence alone to be considered” (ME 112). This is contrasted with the notion of ultimate good unqualified by reference to a particular subject. This unqualified usage of ‘good’ is said to mean “what as a rational being I should desire and seek to realize, assuming myself to have equal concern for *all* existence” (ibid.) Sidgwick goes on to claim that this latter notion of ‘good’ is what we have in mind when we judge conduct to be intrinsically good (ibid.). As I explained in chapter 2, Sidgwick treats the notions of ‘intrinsic good’, ‘absolute good’, ‘objective good’ and ‘universal good’ as equivalent.¹ Recall also the point that Sidgwick’s use of the phrase ‘equal concern for *all* existence’ in the characterization of universal good is misleading in that it gives the impression that if universal good exists, each individual’s welfare necessarily constitutes an equal portion of it. Because there is no contradiction in supposing that one person’s welfare is more intrinsically valuable than another’s, it might have been better for Sidgwick to describe universal good as that which all rational beings as such should desire and seek to realize, regardless of how they are related to it. This characterization more clearly expresses the fact that the universal point of view involves an impartial starting point, even if it might end up requiring

¹ See p. 42 n. 16.

greater concern for some parts of existence over others should we discover that those parts contain more intrinsic value. When I adopt the point of view of the universe I imagine that I have no connection or relation to anyone or anything in particular, including myself. Anything that I ought to desire from this neutral starting point is part of universal good. But it is not a conceptual truth that the good of each individual will constitute an equal portion of universal good--that is a substantive proposition expressed in the axiom of Personal Irrelevance.

The key question for present purposes is whether we have adequate grounds for believing that there is any such thing as universal good. The egoist is able to resist Sidgwick's deduction of impartialism because she refuses to acknowledge that universal good exists. Proponents of egoism maintain that when the individual brackets her unique relation to herself and the special concern she has for her own welfare in order to take up the point of view of the universe, there is nothing that she ought to desire. Egoists maintain that there is only agent-relative good, and practical reasons can only be sourced in facts about oneself such as the fact that a particular course of action will promote one's own happiness.

Despite the egoist's refusal to acknowledge universal good, Sidgwick claims that belief in the existence of universal good is part of the common moral consciousness of mankind. In responding to a critique of his treatment of egoism in the *Methods*, Sidgwick draws our attention to the common belief that the universe was created by an intelligent designer for the purpose of realizing good. He notes that in this context the notion of 'good' must be used 'absolutely', that is, universally (EEM 27-28). The thought here is that the type of good that would serve as God's purpose in creating the world would not be relative to each individual, but rather something that all rational beings as such would have reason to desire regardless of how they are related to it. In his discussion of the 'Proof of Utilitarianism' in the *Methods*, Sidgwick even suggests that an

egoist can be moved to accept impartialism if he acknowledges that his own existence is the product of intelligent design, and therefore part of universal good (ME 421).² Sidgwick notes, however, that the egoist can simply refuse to accept that his existence is part of a divine plan. And the widespread assent to the idea of universal good is insufficient to demonstrate that the egoist is mistaken in believing that the only reasonable ultimate end is the promotion of self-interest.

2.2 *Non-conflict Cases*

Sidgwick admits that he finds egoism to be a somewhat unattractive and objectionable view. He notes that when considered in the abstract, egoism gives offense to “our sympathetic and social nature” (ME 199). He adds further that one’s own happiness is an end that is seemingly ill-suited to be the ultimate end of rational conduct because it is “so narrow and limited, of such necessarily brief duration, and so shifting and insecure while it lasts” (ibid. 403).³ Nevertheless, Sidgwick does not see adequate grounds for rejecting egoism altogether. Not only is it reasonable for the egoist to deny one of the main sources of belief in universal good—that the world is the product of divine decree—Sidgwick believes that the egoist’s position has independent intuitive support. Note first the intuitive plausibility of the thought that the promotion of one’s own happiness is not an optional end contingent on one’s desires but is rather a ‘manifest obligation’ (ME 7).⁴ Sidgwick also finds it intuitive that it would be irrational to

² Sidgwick does not explicitly mention God or intelligent design in the relevant passage. Rather, he uses the phrase ‘designed by nature’. It is not entirely clear why Sidgwick does not mention God explicitly here, but given the context it is clear that this is the idea he has in mind. If the egoist’s acknowledgement that he was ‘designed by nature’ meant only that he was the product of evolutionary forces, it is unclear how this could be seen as an admission that his own good is part of universal good.

³ See Crisp (2015: 195-99).

⁴ On this point Sidgwick references Butler’s *Sermons on Human Nature*.

sacrifice one's own happiness for the sake of that of another (ME 404 n. 1; 498).⁵ Indeed, in recounting his dissatisfaction with Mill's defense of utilitarianism, Sidgwick emphasizes the lack of explanation for why self-sacrifice is ultimately reasonable. Mill's suggestion that such sacrifice would be 'heroic' is not enough: "I must somehow *see* that it was right for me to sacrifice my happiness for the good of the whole of which I am part" (ME xviii).

The fact that Sidgwick frequently appeals to the idea of self-sacrifice when describing the intuitive basis of egoism is significant. If the dialectical situation were such that we had to decide between egoism and impartialism based solely on the *prima facie* plausibility of the views and their respective implications, intuitive judgements about those implications would be highly relevant. However, Sidgwick's attempted axiomatic deduction of impartialism makes the dialectical situation different in an important respect. Given the apparent self-evidence of the axioms of Equal Good and Personal Irrelevance, all that is needed to undermine the egoist's position is sufficient reason for believing that universal good exists. And making a strong case for universal good does not require directly challenging the view that it is always irrational to sacrifice one's own happiness. It is only necessary to provide compelling evidence that there is at least one end such that all rational beings have at least some reason to promote it, regardless of how they are related to it. I believe such evidence can be found by considering cases in which an individual can make substantial contributions to the welfare of others without having to sacrifice any part of her own welfare. The following is one such case:

⁵ Hastings Rashdall (1907: 57) suggests that Sidgwick found egoism more plausible than he should have as a result of the terminology he uses. Sidgwick often uses the terms 'reasonable' and 'rational', which can be taken to indicate internal inconsistency or means-ends coherence. I shall return to this point in chapter 4.

Faraway People: In a faraway galaxy there exists a planet filled with persons very similar to human beings. One day, an individual named Mohatten is presented with an opportunity to make a drastic difference in the lives of the distant strangers. Mohatten is presented with two buttons—one green, and one red. He is informed that if he presses the red button nothing will happen to the faraway people, and if he presses the green button, each of the faraway people will receive an extra ten years of happy life. If he presses either button, Mohatten will receive an extra ten years of happy life. If he does not press either button, nothing will happen to him. Mohatten does not care at all about the faraway people, and he knows he will never have any contact with them. Mohatten is informed that no matter which option he chooses, nobody will know his decision, and he will immediately be made to forget about the entire scenario.

The egoist is committed to the claim that Mohatten would have just as much reason to press the red button as he would to press the green button. This is because the egoist denies that the welfare of others is an independent source of reasons.⁶ Though the happiness of others is certainly good for them, the egoist maintains that each individual ought only to aim at his or her own good. However, it seems quite clear that anyone faced with such circumstances would have decisive reason to press the green button. Proponents of egoism and impartialism would agree that Mohatten should press at least one of the two buttons in order to obtain the extra ten years of happy life. But the fact that pressing the green button will confer substantial benefits to many

⁶ This is not true of every conceivable version of egoism. A person who thinks that the good of others is a source of reasons but thinks that self-interest always trumps the good of others in cases of conflict qualifies as an egoist. But that is not the sort of egoist we are concerned with for present purposes. The chief threat to Sidgwick's deduction of impartialism is the view that there is only agent-relative good. Hence, we are concerned only with those egoists who deny that the interests of others are an independent source of reasons. Thanks to Jason Gardener for urging me to clarify this point.

people makes it hard to see how there could be just as much counting in favor of pressing the red button. And because Mohatten has no antecedent connection to the faraway people, and because he has no desires concerning them, the presence of a reason to press the green button rather than the red suggests that the welfare of those people is an end that one ought to desire and seek to promote regardless of how one is related to it. In other words, the welfare of others is part of universal good. One might object that pressing the green button is supported by prudence because of potential guilty feelings that might arise after pressing the red one, or because the harsh judgement one would receive from others. However, this worry is precluded by the fact that immediately after pressing the button Mohatten will be made to forget that he was ever presented with this decision, and nobody else will know which choice he has made.⁷

There are possible lines of response one might make in order to defend the claim that there is equal reason to press either button. A proponent of egoism would likely appeal to the metaphysical distinction between persons to explain her judgement about the case. Recall from the previous chapter that Sidgwick utilizes this idea in order to explain why the egoist's position is difficult to undermine:

It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently "I" am concerned with the quality of my existence as in individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual. (ME 498).

⁷ One might worry that appealing to intuitions about hypothetical cases is inconsistent with Sidgwick's critique of perceptual intuitionism. But Sidgwick is not principally opposed to the use of thought-experiments and imaginary cases as part of ethical inquiry. What Sidgwick disparages is the notion that one's own conscience is the ultimate arbiter of ethical truth, and that basic ethical principles are not necessary.

The suggestion that the separateness of persons is of high ethical importance is not unique to Sidgwick. John Findlay calls it ‘the fundamental fact for ethics’, and Rawls’ primary criticism of utilitarianism is that it fails give adequate weight to the distinction between persons.⁸ Hence, it does not seem unreasonable for egoists to appeal to the separateness of persons in an attempt to justify their judgement about the ‘faraway people’ example. The egoist could argue that the explanation for one’s having equal reason to press either button is that the faraway strangers are distinct persons, and that the separateness of persons plays a fundamental role in determining one’s practical reasons.

I readily grant the initial plausibility of the thought that the fact that I am connected to my own consciousness in a way in which I am not connected to others has non-derivative ethical significance.⁹ Perhaps the fact that I am me and you are you would give me a reason to save my own life rather than yours if forced to choose between the two. However, in order for the separateness of persons to do the work that the egoist needs it to do, it is not enough for the distinction to have *some* degree of significance. In order to avoid the conclusion that there is decisive reason to press the green button in ‘faraway people’ the egoist needs it to be the case that the significance of the distinction between individuals is absolute. The significance must be so great that one could have decisive ultimate reason to promote one’s own happiness, but absolutely *no* ultimate reason to promote the qualitatively identical happiness of another person. While this suggestion is not incoherent, it is highly implausible. C.D. Broad (1930: 245) puts the point forcefully in his assessment of Sidgwick’s discussion of egoism: “It does seem to me conceivable, though not self-evident, that I ought to desire *more strongly* the occurrence of a good state of mind in myself than the occurrence of an equally good state of mind in anyone else;

⁸ Findlay (1961: 294); Rawls (1971: 27). References from Parfit (2011, 1: 498).

⁹ That being said, I shall challenge this idea in chapter 4.

whilst it seems self-evident that I ought to desire *to some degree* its occurrence anywhere.” In other words, while it might be true that the separateness of persons matters for ethics, if it does it is only a matter of degree.

An alternative way in which someone might try to defend the egoist’s judgement about ‘faraway people’ is to appeal to facts about the individual’s desires. One might try claiming that if pressing the green button rather than the red one would not satisfy any of the agent’s desires, this explains why there is equal reason to press either one. One could also appeal to metaethical concerns about the very notion of something being good in itself such that all rational beings have reason to promote it regardless of whether they care about it or how they are related to it. It might be said that universal good is metaphysically queer, or that we would have no reliable means of recognizing truths about such good.

These replies, however, cannot succeed. Recall from chapter 1 that we are working under the assumption that subjectivism is false, and that practical reasons are object-given and value-based. This assumption is shared by Sidgwick in the *Methods*, and the brand of egoism he is troubled by holds that the promotion of self-interest is an unconditional requirement of reason. Sidgwick’s egoist believes that one ought to promote one’s own good regardless of whether doing so would satisfy one’s desires. Moreover, since this brand of egoism implies that practical reasons are objective rather than subjective, any metaphysical and epistemological worries one might have about universal good apply equally to the reasons sourced in one’s own good.

Not only does the commitment to objective normativity preclude egoists from appealing to facts about desires in order to defend their judgement about ‘faraway people’, it also highlights the oddity of the egoist’s overall position. It is easy to see why proponents of subjectivism would deny the existence of universal good. As we saw in chapter 1, subjective

theories hold that practical rationality is ultimately a matter of getting what one wants. Even in their most sophisticated forms such as Bernard Williams' idealized theory (1981) and Sharon Street's Formal Humean Constructivism (2012), these views imply that an agent has reason to Φ only if she genuinely cares about Φ -ing. If practical reasons were grounded in facts about the agent's desires, it would be clear that the locus of normativity is the individual herself, and it would be clear that there could not be an end that all rational beings as such have reason to promote. Indeed, Parfit suggests that one possible explanation for someone's having a completely egoistic outlook is that she has been taught to accept some desire-based subjective theory of practical reason (2011, 1: 137).

But given the assumption that practical reasons are object-given and value-based, the thought that there is only agent-relative good looks strange. To appreciate the strangeness of this view, consider the example of Anita. Anita is suffering from an illness that is potentially life-threatening. Fortunately, a new medicine has become available that is highly effective in treating this illness. If she takes the medicine, Anita will be cured and go on to enjoy many years of happy life. According to egoists, Anita has decisive reason to take the medicine regardless of whether doing so will satisfy any of her desires (whether actual or idealized). This point by itself is not difficult to accept. But what is strange is the further implication of the view which is that nobody else could have *any* reason to try to help Anita obtain the medicine unless doing so would promote their own interests. If Anita has reason to take the medicine regardless of whether her doing so would satisfy any of her desires, why wouldn't it be the case that other people have at least *some* reason to help her acquire the medicine regardless of how this would affect them? It would be one thing if the grounds for Anita's reason to take the medicine were the fact that she wants to live, and that taking the medicine would save her life. We could then

say that the reason others do not have reasons to help her is that they do not care whether she lives or dies. But we are working under the assumption that practical reasons are grounded in objective features of objects and outcomes rather than desires or preferences. If the objective features of Anita's future happy life are part of the ground of her reasons to take the medicine, it is hard to see why these features could not give others at least *some* reason to help her. And if other agents would have at least some reason to help Anita regardless of whether doing so would promote their own welfare, this means that Anita's welfare is part of universal good.

3 The Value of Partiality

I take it that the foregoing discussion provides adequate grounds for accepting the existence of universal good. But even if we could remove all doubt about the existence of universal good, this would not be sufficient for a successful deduction of impartialism from Sidgwick's axioms. The acceptance of the axioms of Personal Irrelevance and Equal Good is consistent with the rejection of Sidgwick's principle of Benevolence (which is equivalent to the view I call Rational Impartialism). A proponent of partiality might agree that everyone's welfare constitutes an equal portion of the good and that practical reason precludes giving unequal weight to equal portions of the good, while also maintaining that there are ultimate reasons of partiality nonetheless. These ideas would be consistent with one another if partiality (instantiated in relationships themselves, acts of love, displays of loyalty, etc.) is valuable in and of itself. If love-based partiality is valuable for its own sake, then there could be circumstances in which promoting the lesser needs of a friend or family member rather than the greater needs of a stranger could be justified by the light of reason. This would be the case whenever the value realized through the

display of partiality is greater in quantity than the value inhering in the welfare that was forsaken by prioritizing the lesser needs of a loved one over the greater needs of a stranger.

Suppose, for example, that you face a choice between volunteering at an under-staffed homeless shelter, and spending the same amount of time relaxing at home with your significant other. Spending the extra time with your partner will certainly promote the welfare of both of you, but we can suppose that you will make a more substantial contribution to general welfare by volunteering instead. Even if this is right, it could still be the case that more overall good would be realized by choosing your partner over the shelter. This could be true if displays of love or relationships themselves have their own inherent value—a value above and beyond the value inhering in their contribution to welfare.

In light of these considerations, it is clear that Sidgwick's axioms cannot rule out ultimate reasons of partiality as long as we are justified in believing that partiality is itself a source of intrinsic value. This is why my Sidgwickian argument for Rational Impartialism includes a premise stating that partiality towards one's intimates is only derivatively good (premise 5). The rest of this chapter is dedicated to defending this premise.¹⁰

3.1 Personal Relationships

I suspect many would agree that the value of acting for the sake of loved ones can be greater than the value realized by benefitting a stranger, even if the benefit to the stranger would be more substantial. For instance, if I must choose between buying a gift for a close friend or lover and

¹⁰ Sidgwick does not directly address this issue in the *Methods*. In the chapter in which he argues for hedonism, the alternative candidate sources of intrinsic value he considers are virtue, truth, beauty, and freedom (ME 391-407). The most thorough discussion of personal relationships occurs in the context of an examination of common sense morality. Sidgwick argues that, as with the majority of common sense morality, the purported duties of partiality towards intimates lack the clarity, precision, and unification necessary for practical as well as theoretical purposes (ME 244-263).

donating that money to famine relief, many of us would judge that I ought to buy the gift. We make this judgment despite the fact that the donation could save a stranger's life, and my friend or lover already has a wealth of material possessions. Gift giving is just one of many ways in which we exhibit partiality towards intimates. Another example is our tendency to devote time to meeting their needs even if we could do more overall good by using that time helping strangers (perhaps by volunteering or campaigning). Ostensibly, the reason for choosing to buy the gifts and spend time helping our loved ones even when our time and money can do more good for strangers is that prioritizing the good of our loved ones is an essential component of intimate relationships, and such relationships are highly valuable for their own sake.

Though the idea that close personal relationships are intrinsically valuable is initially plausible, it does not withstand critical scrutiny. The primary reason for doubting that personal relationships are valuable for their own sake is that they fail isolation tests for intrinsic value.¹¹ If relationships really are valuable for their own sake, we should find that there is at least some reason to preserve or promote them that is independent of their contribution to other sources of value such as hedonic welfare. But, as I shall now try to show, this is not what we find.

Imagine, for example, that we discover an isolated culture in a heretofore unexplored territory.¹² We observe that the people in this society are completely impartial in their dealings with others, and that they do not have anything approximating the close personal relationships that we are accustomed to. They do not have monogamous romantic relationships or close friendships, and their children are raised communally. Suppose that the social structure of this

¹¹ Sidgwick applies an isolation test of sorts in considering whether 'ideal goods' such as virtue, truth, beauty, and freedom are intrinsically valuable. He claims that when we isolate these goods from the positive consciousness accompanying them, they do not appear to be desirable (ME 400-1). G.E. Moore's (1903: 83-5) 'two worlds' thought experiment is perhaps the most influential example of an isolation test for intrinsic value.

¹² The following example (as well as the example in the next paragraph) is a modified version of an example found in Simon Keller's *Partiality* (2013).

society works very well insofar as the individuals living there enjoy high levels of happiness, as well as intellectual, artistic, and athletic achievement. If close personal relationships were valuable for their own sake, there would be reason to lament the fact that the people living in the society are completely impartial. We would have reasons to hope that they would rearrange their social structure, and perhaps we might even have reasons to try to help them make the change. But given that the people in this society are every bit as creative, productive, and content as we could hope to be, it is hardly plausible that we would have any such reasons.

Another way to see the same point is to consider the example of a romantic relationship between two individuals who, after being together for several years, have slowly grown apart. Because of an increasing divergence in their goals and desires they often quarrel, and they no longer enjoy each other's company. Once it becomes clear that preserving the relationship will not contribute to the flourishing of either party there is no reason for them to try to maintain it. It would be strange to suggest that they should stay together to their own detriment *for the sake of the relationship itself*.¹³ Similar remarks apply to the case of friendships.

Not only does the apparent value of friendships and romantic partnerships seem to disappear when we isolate them from their contributions to welfare, it is also the case that we tend to assign value to such relationships roughly in proportion to the amount of welfare that they bring to the individuals involved.¹⁴ When two individuals are able to reach great levels of flourishing within a relationship, we judge the relationship to be of especially high value. When the benefit to the individuals is not as great, we assign less value to the relationship.

Here one might object that it is not just any close relationship that has intrinsic value, but only healthy ones. Of course a relationship that is abusive or otherwise dysfunctional is of no

¹³ See Keller (2013: 56-58).

¹⁴ Sidgwick makes a similar point regarding common judgements about the value of knowledge, beauty, truth, freedom, and virtue (ME 401).

value, but this is compatible with healthy relationships having value in and of themselves.

Perhaps this intrinsic value can only arise under conditions in which the relationship contributes to the flourishing of both parties. Those who claim that relationships have intrinsic value might even insist that their use of the term ‘relationship’ is meant only to refer to healthy associations between individuals.

To see why this objection fails, it will help to consider the following analogy provided by Simon Keller:

I claim that winter coats are intrinsically valuable, meaning that they have value in addition to the good they do in keeping people warm. You say that it seems implausible to think that a winter coat could be valuable even when it does nothing to keep anyone warm. I reply that it wouldn't be; I say that the intrinsic value of a winter coat is conditional upon the coat's also having the value of keeping someone warm, or perhaps that a winter coat that does not keep anyone warm is not really a “winter coat” in the sense in which I use the term. You would be within your rights to think that I am just trying to make trouble. Given my concessions, what more could I possibly need to see that the value of a winter coat is purely extrinsic? (2013: 58).

The point of this analogy is that acknowledging that contribution to welfare is a precondition for a relationship's being valuable leaves us without any philosophical motivation for claiming that relationships are valuable for their own sake.¹⁵

None of this is to deny that loving relationships are highly valuable, and that in practice we often have reasons to show partiality towards friends, family, and romantic partners. But the above considerations suggest that the true value of loving relationships is found in their

¹⁵ A similar point can also be used to defend Sidgwick against a criticism offered by Crisp (2015: 82-3). In discussing the purported intrinsic value of ‘ideal goods’ such as truth and beauty, Sidgwick argues that when we isolate these goods from the consciousness accompanying and resulting from them they do not appear to be of any value (ME 400-1). Crisp objects that even though an individual's knowledge may appear valueless in isolation, this does not imply that such knowledge combined with the conscious states is of no greater value than the conscious states themselves. Crisp is certainly right about this. However, if the ‘ideal goods’ appear valueless in isolation from the positive experiences they give rise to, there is little motivation for believing that they have independent value of their own over and above those experiences.

contribution to the welfare of the individuals who participate in them. Importantly, if our reasons to preserve and promote relationships depend on matters of welfare, this suggests that our reasons *within* relationships are dependent in the same way. Whatever reason we have to pay special attention to the needs of those near and dear to us derives from the fact that our relationship puts us in a unique position to promote their interests. A shared history and common values allow me to provide emotional support to my friend that I could not provide to just anyone.¹⁶ I have reason to cancel my weekend plans to volunteer at the homeless shelter in order to support my grief-stricken friend at his wife's funeral not because friendships are valuable for their own sake, but rather because my friend's welfare matters, and I am in a unique position to make a substantial contribution to it. But of course, my friend's welfare is no more important than a stranger's. So while our relationship gives me reasons to treat my friend in many ways differently from the way I treat strangers, it does not give me reasons to act so as to bring about a state of affairs that is worse overall but better for my friend.

Since much of the preceding discussion has drawn on the arguments Keller presents in *Partiality* (2013), it is worth considering his suggested rationale for fundamental partiality towards loved ones. Though he denies that loving relationships are valuable for their own sake, Keller maintains that we have ultimate reasons of partiality nonetheless. While he acknowledges the difficulty in finding a philosophical rationale for the existence of ultimate reasons of partiality, he suggests that the phenomenology of our ethical experiences might be sufficient. Keller claims that finding another person to matter for her own sake "is one of our deepest and most familiar ethical experiences" (2013: 56). He adds further that, "It is that experience, not

¹⁶ Robert Goodin (1985: 33) helpfully characterizes reasons of partiality as reasons to protect and advance another person's interests in contexts in which that person's interests are especially vulnerable to our actions. This is typically the case with one's friends and family.

just a brute intuition, that the belief in reasons of partiality is grounded, and it would take quite some argument to show that the experience rests on an illusion” (ibid.).

Denying that there are ultimate reasons of partiality does not require showing that the ethical experience Keller refers to is an illusion. Impartialist theories of practical reason do not deny that individuals are valuable for their own sake. What the impartialist denies is that recognizing a person’s intrinsic value justifies rationally discounting the value of relevantly similar strangers. When I see my best friend, I have the deep and familiar ethical experience Keller alludes to—I see someone who matters in her own right. But the reason she matters is that she has capacities for things like enjoyment, suffering, and achievement. People whom I have never met (and may never meet) matter just as much because they have these same capacities. Admittedly, I am motivated to care for my friend more than I am to care for distant strangers—she matters more *to me*. What I take to be illusory is the common thought that this fact about my motivations and evaluative attitudes reveals something deep about my ultimate practical reasons.

It also bears mentioning that Rational Impartialism is compatible with the intrinsic value of things such as compassion and kindness. If it turns out that acts of compassion or kindness towards other beings have value over and above the promotion of welfare, this does not undermine the claim that our ultimate practical reasons are impartial. This is because compassion and kindness can be instantiated towards strangers just as well as it can towards one’s intimates.

3.2 Love

The preceding discussion focused primarily on the question of whether relationships themselves are a source of intrinsic value. Some may be unmoved by these considerations because they believe that it is not relationships that are the source of value but rather love itself. It is thus worth considering whether love is a source of value over and above the contributions it makes to welfare. Though love is certainly not easy to define, for present purposes we can consider a plausible characterization offered by Sidgwick in the *Methods*. According to Sidgwick, love is primarily a pleasant emotion comprising three essential components: (1) a sense of union with another person, (2) a desire to do good to the beloved, and (3) a desire for the society of the beloved (ME 244). Sidgwick considers whether love so understood should qualify as a moral excellence. Though he thinks it is natural and desirable that we be moved by affection for particular individuals, this is only so to the extent that such affection can help us fulfill particular duties we might have, and he denies that we should cultivate love beyond this point (ME 245).

Our question about love is different from Sidgwick's question. We are not asking whether love is a moral excellence but rather whether it is a source of intrinsic value capable of generating reasons that can override reasons grounded in welfare. To answer this question we can examine the three components of love listed above. Consider first the sense of union with another person that is characteristic of love. It is highly plausible that this sense of union is valuable. But as with the case of relationships themselves, the value seems to be located in the contribution to a person's wellbeing. Given the sorts of creatures we are, a sense of union with another person is typically a great source of contentment, enjoyment, and security. And these three products of a loving union undoubtedly make a person's life go better. But if we consider a person who is able to attain the same levels of happiness and gratification without being in love,

it does not seem like she would still have reason to try to fall in love. And when we imagine two people who share a sense of union with each other, but do not gain any of the prudential goods that typically arise from this, it does not seem that there is anything of value there. In other words, a sense of union with another person does not appear to have value independent of the contributions it typically makes to welfare.

The second element of love as characterized by Sidgwick is a desire to do good to the beloved. It is fairly clear that a benevolent impulse towards others is valuable insofar as we can often promote the good of others in ways that they themselves are unable to, such as when we express sympathetic concern for a friend experiencing grief. But setting this instrumental value aside, it tempting to think that there is a special value inhering in the benevolent desire itself. But even if this were the case, it is doubtful that the value of the benevolent impulse is limited to benevolence towards specific individuals. Consider an agent who has a strong desire to promote general welfare, and so she sets out to help those who are worst off, regardless of where they are located or how she is related to them. Compare this with the example of an agent who feels an equally strong impulse to promote the welfare of her lover. Perhaps both of these impulses are valuable in their own right, but it is hard to believe that the later agent's desire is *more* valuable. And if the impulse to aim at good generally is just as valuable as the impulse to aim at a particular part of it, this undermines the suggestion that the value of love is a source of ultimate reasons of partiality.

The third element of love is the least plausible source of intrinsic value. Viewed in isolation, the desire for the society of the beloved does not seem valuable even instrumentally. The bare desire for companionship with another is no more plausible a candidate for intrinsic value than other biological desires such as the desire for food, shelter, and rest. Moreover, the

desire for companionship is often a source of disvalue. As Sidgwick observes: “this element may predominate over [the desire to do good to the beloved], and even conflict with it, so that the true interests of the beloved may be sacrificed. In this case we call the affection selfish, and do not praise it at all, but rather blame” (ME 244-45). That being said, when two individuals are in a loving relationship their knowledge of each other’s desire for communion can contribute to their welfare. But this awareness is essentially the sense of union already discussed above.

3.3 Parental Partiality

Intuitions supporting the intrinsic value of partiality are perhaps strongest in the case of parents and their children. While common sense acknowledges a basic obligation to nurture and protect one’s children, it also supports the idea that the special care and attention most parents provide for their children has value in and of itself. But as is the case with friendships and romantic relationships, the purported intrinsic value parental partiality does not survive critical scrutiny.

One need not conduct a survey of psychological literature to know that human children benefit from experiencing a close bond with their parents. The love and support of one’s parents provides a sense of security for a child, and helps them to develop intellectually and emotionally. Further, for many (if not most) parents, the experience of intense love for their children provides more happiness and contentment than anything else they experience during their lives. In light of these points, it is clear that parental partiality is highly important. But imagine a scenario in which partiality towards one’s children would not result in an improvement of the welfare of the child or the parent. The earlier example of a society in which children are raised communally will suffice for this purpose. Suppose once again that the people that constitute this society experience just as much happiness and prosperity as those living in societies in which children

are raised by parents who prioritize their needs over the needs of others. It does not seem to be the case that the absence of parental partiality in this society is something to lament. Of course, the individuals living there miss out on the wonderful enjoyment and sense of purpose that many of us experience through the special bonds we have with our parents and children. But if they find just as much meaning and enjoyment in their lives without parent-child bonds, then this is not a genuine loss.

Though Sidgwick does not consider whether partiality towards one's child is valuable for its own sake, he does discuss the purported duties of parents to their children, and his comments are instructive for both issues. On the question of whether parents have basic duties to provide for their children Sidgwick writes:

We have no doubt about this duty as a part of the present order of society, by which the due growth and training of the rising generation is distributed among adults... If, however, we consider the duty of parents by itself, out of connexion with this social order, it is certainly not self-evident that we owe more to our own children than to others whose happiness equally depends on our exertions. To get the question clear, let us suppose that I am thrown with my family upon a desert island, where I find an abandoned orphan. Is it evident that I am less bound to provide this child, as far as lies in my power, with the means of subsistence, than I am to provide for my own children? (ME 346-7)

Although we tend to judge that a parent's love and dedication to her own child has value in and of itself, Sidgwick's thought experiment helps us to see that this is largely due to a failure to consider contexts in which just as much good would be realized by giving equal weight to the good of someone else. Although I certainly have reasons to try to ensure that my child does not suffer and has a minimally decent life, this is best explained by facts about the badness of suffering and the positive value of enjoyment, achievement, autonomy, etc., combined with the fact that I am in a unique practical position with respect to the interests of my child. I am in this

unique position primarily due to contingent facts about the social structure of the society in which I live. If I happen to be in a unique position to care for an unrelated child, I would have equal reason to do so.

Given our strong biological disposition to care for our young, it should not be surprising that we tend to view parental partiality as a manifest obligation. These facts about our motivational psychology and the social structure of society also explain why we are prone to judge partiality towards children to be valuable for its own sake. But on “Sidgwick’s Island” I would have just as much reason to care for the orphaned child as I would for my biological children. This is because the orphaned child has all the same ethically relevant capacities as my own child, and she is equally lacking in ability to care for herself. The fact that I am not her biological parent is irrelevant.

It is highly plausible that it is better for the world as a whole if parents are disposed to favor their own children because the absence of such partiality would result in an entire generation of people who develop emotional and psychological deficits of varying degrees. But this suggests only that partiality towards one’s children is derivatively good. I ought to pay special attention to the needs of my child because this is what is best from an impartial perspective. It is ultimately a good thing that my motivational psychology is such that it would be extraordinarily difficult for me to save the lives of several unrelated children at the cost of letting my own child die. But it does not follow that I would have an ultimate reason to save my child.¹⁷ If the grounds for partiality toward one’s children is the impact on the general welfare of everyone, we cannot have ultimate reason to bring about a state of affairs with less total welfare merely to protect that of our own children.

¹⁷ The lack of an ultimate reason to save my child rather than several unrelated children is compatible with my having ultimate reasons to deliberately maintain and cultivate the sentiments and inclinations that would likely move me to save my child’s life under such circumstances.

3.4 Partiality and Self-Sacrifice

One explanation for the intuition that partiality towards loved ones is intrinsically valuable is the fact that such partiality often involves some degree of self-sacrifice by the agent in question. Of course, devoting time and other resources to our loved ones is an excellent way to promote our own happiness. But loving relationships also require us to set aside our own plans and projects in order to promote the welfare of the beloved. The willingness to sacrifice one's own interests for the benefit of one's intimates certainly commends itself to common moral consciousness. But while this might be taken as evidence that there is intrinsic value in acts of sacrifice for the sake of loved ones, there are a few important points that undermine this thought.

First, just as we feel admiration for those who sacrifice themselves for the sake of friends and family, we also feel admiration for those who sacrifice themselves for the sake of strangers. Even if stories of Good Samaritans do not have the same emotional impact as stories of sacrifices made for loved ones, we are certainly moved when we hear of individuals willing to set aside their own good for the sake of another human being whom they have never met. This suggests that what we are responding to when we feel great admiration for those who sacrifice themselves is not the fact that they are showing partiality to the person they are helping, but rather the fact that they are capable of overcoming their self-interested instincts for the sake of others. The special feeling of approval in either case can be attributed, at least in part, to the fact that we are predisposed to self-regard, and a willingness to suppress this is good from an impartial perspective.¹⁸

Admittedly, there are cases in which our responses to acts of sacrifice for the sake of others differ depending on whether the recipient is a loved one or a stranger. For instance, while

¹⁸ Of course, if Rational Impartialism is true, one could not have reasons to sacrifice one's greater interests for the sake of the lesser interests of others. I discuss this point in further detail in chapter 5.

kidney donations made for the benefit of a friend or family member are typically the subject of universal admiration, not everyone responds favorably to cases of individuals who donate kidneys to strangers. Not only do some people respond more favorably to the former sort of case, there are some who actually have negative reactions to the latter.¹⁹ But once again, we need not assume that these divergent reactions are tracking any deep ethical truths. The divergence is well explained by the fact that we are biologically predisposed to feel strong inclinations to protect our loved ones, combined with the fact that voluntary organ removal runs counter to our typical notions about health and the role of medicine. While our disposition to protect and provide for our intimates is strong enough to overcome our aversion to organ removal such that we approve in the case of donations for family and friends, our inclinations to help whoever happens to be in need are not always strong enough to overcome the aversion. Further, it is plausible that the negative reaction to those who donate kidneys to strangers arises from feelings of guilt for not being willing to make the same sacrifice (ibid.)²⁰

4 Conclusion

In this chapter, I have presented the case for accepting the existence of universal good. I have argued that the existence of universal good is most readily seen when we consider cases in which promoting the welfare of strangers does not require any sacrifice on the part of the individual. I have also argued that the denial of universal good sits uncomfortably with the egoist's commitment to an object-given account of practical reasons. The vindication of universal good

¹⁹ Andrew Anthony, "Would You Give Your Kidney to a Complete Stranger," *The Guardian* (July 2014): <http://www.theguardian.com/society/2014/jul/13/kidney-donation-altruism-andrew-anthony>

²⁰ Ibid.

means that the axiom of Equal Good should be read with a universal specification, and that premise 2 of my Sidgwickian argument for Rational Impartialism is secured.

The second task of this chapter was to defend premise 5, which states that partiality towards one's intimates is only derivatively good. I applied isolation tests of intrinsic value to undermine the common thought that loving relationships and the acts of partiality they give rise to are valuable for their own sake. Though we clearly have reasons to pay special attention to those near and dear to us, these reasons are explained by the unique practical and epistemic position we are in relative to their needs.

Had Sidgwick been confident in the existence of universal good, his masterpiece would not have ended on the somber note that there is "an ultimate and fundamental contradiction in our apparent intuitions of what is Reasonable in conduct" (ME 508). He would have concluded that the axioms of Equal Good and Personal Irrelevance are sufficient for vindicating an impartialist account of practical reasons. However, even if the arguments presented thus far are successful, this is not in fact enough to secure Rational Impartialism. This is because the self-evidence of the axiom of Equal Good is called into question by proponents of the view I call Partial Dualism. It is the task of the next chapter to examine the motivations for rejecting Equal Good, and to present reasons for believing that such rejection is a mistake.

4 'Equal Good' and the Partial Dualism of Practical Reason

1 Introduction

Recall that my argument for Rational Impartialism takes the following form:

1. One ought to have equal fundamental concern for equal portions of the good.
[Axiom of Equal Good]
2. The good is not relative to individuals but rather is universal.
3. One ought to have equal fundamental concern for equal portions of universal good.
(from 1-2)
4. The equal good of any two individuals is exactly similar qua portions of universal good. [Axiom of Personal Irrelevance]
5. Partiality towards one's intimates is only derivatively good.
6. One ought to have equal fundamental concern for the equal good of all individuals.
(from 3-5)

In the previous chapter I argued that we have strong grounds for believing that the good is universal rather than agent-relative. I also provided reasons for thinking that the value of partiality is merely derivative. If these arguments are successful, then premises 2, and 5 of my argument are secured. Given that the premises 1 and 4 are two of Sidgwick's axioms, it might appear that the Sidgwickian case for Rational Impartialism is complete. However, as I noted previously, the axiom of Equal Good is more contentious than Sidgwick realizes.

Sidgwick is aware of the fact that some common-sense moralists will reject his axiom due to their belief in special duties of partiality that are a source of ultimate reasons (e.g. parental duties, duties of friendship, duties of gratitude, etc.). Sidgwick does not find this disagreement over Equal Good worrisome because he sees several reasons for believing that common-sense is

mistaken on this score.²¹ What Sidgwick does not anticipate, however, is that some of his most capable successors—philosophers who share many of his methodological and ethical views—implicitly reject the axiom of Equal Good. Roger Crisp (2006) and Derek Parfit (2011) each champion the view I call Partial Dualism. On this view, the welfare of all is a source of reasons for all, but one’s own good (and for Parfit, the good of one’s intimates) is a source of reasons with independent weight over and above the weight they would have in an impartial calculus.²² Partial Dualists thus deny that one ought to have equal fundamental concern for equal portions of the good because they believe individuals are justified in prioritizing their own good over the equal good of strangers. The explanation for this purported rational disparity between self and others is that value is not the only source of an agent’s reasons—it also matters how the agent is related to that value. The connections I have to my future self is claimed to be a non-evaluative source of reasons that can compete with reasons sourced in the universal value of everyone’s welfare.

In light of the Partial Dualist’s challenge, my argument for Rational Impartialism cannot succeed without a compelling defense of Equal Good. To provide such a defense is the primary aim of this chapter. I begin in section 2 with a brief overview of the motivations and commitments of Partial Dualism. In section 3 I articulate the attractions of Equal Good, and I argue that the best strategy for proponents of Partial Dualism is to advocate a qualified version of

²¹ See ME 238-263.

²² Crisp calls his account a “dual-source view”, while Parfit labels his account a “wide value-based objective view.” I have chosen to use the label “Partial Dualism” because it helps to distinguish the views of Crisp and Parfit from other accounts of rational weight such as Sidgwick’s Dualism, Rational Egoism, and Rational Impartialism. Although the views of Crisp and Parfit are similar, it is worth noting two points on which they diverge. First, while Parfit thinks the interests of an agent’s friends and family can also outweigh impartial considerations, Crisp does not. Second, while Crisp’s view implies that personal sacrifice for the greater good can be irrational, Parfit suggests that such sacrifice might always be rationally permissible. The significance of these differences will become clear as the discussion proceeds.

the axiom. In sections 4 and 5 I apply Sidgwick's first two tests for highest certainty to show that we lack adequate justification for abandoning Equal Good in favor of the qualified version.

2 Partial Dualism

Parfit and Crisp are both unabashedly enamored with Sidgwick's arguments and methodology. In the preface to *On What Matters* Parfit calls Sidgwick one of his two masters (the other being Kant), and claims that the *Methods* is "the book that it would be best for everyone interested in ethics to read, remember, and be able to assume that others have read" (2011, 1: xl). Crisp describes the *Methods* as a "philosophical goldmine," and he endorses an earlier philosopher's opinion that Sidgwick's work is a source of "pure white light" (2015: x; 2002: 56).²³

Despite being so greatly influenced by Sidgwick, Parfit and Crisp diverge from their master on a few key issues. The major disagreement concerns Sidgwick's infamous Dualism of Practical Reason. Whereas Sidgwick views self-interest and universal benevolence as equally plausible candidates for the ultimate practical end, Parfit and Crisp aim to show that neither egoism nor impartialism is tenable. While both writers accept that the good of all is a source of reasons for all, they also maintain that one's own good carries greater rational weight for one than the equal good of strangers. Hence, in attempting to avoid Sidgwick's somber view, these contemporary Sidgwickians reject the axiom of Equal Good. My aim in this section is to outline some of the initial attractions and theoretical commitments of Partial Dualism, and to explain how this view constitutes a forceful challenge for Equal Good and my overall case for Rational Impartialism.

²³ The "pure white light" comment was originally made by a British philosopher overheard by Brand Blanshard as recounted in his *Four Reasonable Men: Marcus Aurelius, John Stuart Mill, Ernest Renan, Henry Sidgwick* (1984).

In the *Methods*, Sidgwick writes: “[It] is hardly going too far to say that common sense assumes that ‘interested’ actions, tending to promote the agent’s happiness, are *prima facie* reasonable: and that the *onus probandi* lies with those who maintain that disinterested conduct, as such, is reasonable” (ME 120). Crisp and Parfit each take up the challenge of showing that the interests of strangers can rationally override self-interested reasons. In arguing against egoism, Crisp and Parfit primarily appeal to intuitive judgments about hypothetical scenarios. Consider the following case offered by Crisp:

Two Doors 2: You are confronted by two doors. If you do not pass through one or other of them, you will suffer an extremely painful electric shock. If you pass through door A, you will experience a minor twinge in your leg, but nothing further will happen. If you pass through door B, you will not experience the twinge, but some other person, a stranger and out of sight, will suffer an extremely painful electric shock. (2006: 132)

Crisp suggests that to claim that you have stronger reason to pass through door B rather than door A is almost as implausible as claiming that none of us have any reason to avoid excruciating pain. When we can prevent a stranger’s suffering by making a minuscule self-sacrifice, we have decisive reasons to do so: “The well-being of others, in other words, can ground reasons to act which override the reason we have to promote our own well-being” (ibid.)

In the context of his discussion of the egoism that is implied by Sidgwick’s Dualism, Parfit points out that on this view we could rationally do what we knew would only be slightly better for ourselves but substantially worse for many people: “For example, we could rationally save ourselves from one minute of discomfort rather than saving a million people from death or agony” (2011, 1: 135). After acknowledging that many people might judge such an act rational in the instrumental sense, Parfit asserts that such an act would not be rational in the more important objective reason-based sense: “Would we have sufficient reasons to save ourselves

from mild discomfort, rather than saving a million people from death or agony? The answer, I believe, is No. This horrendous act would not be rational” (ibid.).²⁴

Given their belief that the interests of beings unrelated to oneself are a source of objective, non-derivative practical reasons, one might have expected Parfit and Crisp to endorse an impartialist theory of rational weight. However, both authors find impartialist views implausible. Parfit believes that when the stakes are high enough, such as when one’s life is in danger, giving priority to one’s own interests is rationally justified. He puts the point as follows:

According to Rational Impartialists, we could not have sufficient reasons to do what would be impartially worse than some other possible act. That is not true. We might have such reasons, for example, when and because our act would be much better for ourselves. I would have sufficient reasons to save my own life rather than the lives of several strangers. (2011, 1:131)

Crisp also appeals to intuitions about cases of conflict between self-interest and the interests of others to motivate the rejection of Rational Impartialism. The following is one such case:

Two Doors 3: You are confronted by two doors. If you do not pass through one or the other of them, you will suffer an extremely painful electric shock. If you pass through door A, you will experience a less painful but significant shock. If you pass through door B, you will not experience this shock, but some other person, a stranger and out of sight, will suffer a shock of the same intensity. (2006: 133)

Crisp finds it obvious that you have decisive reason to pass through door B rather than door A because in addition to the reasons provided by the well-being of everyone, you have separate reasons provided by the fact that your own well-being is *yours*.

²⁴ Though he does not provide much by way of explanation for this judgment, Parfit suggest that one reason such acts would not be rational is that they would be morally wrong (2011, 1: 136).

Neither Parfit nor Crisp suggests that there is any precise truth about the relative weight of self-interested and impartial reasons.²⁵ However they do believe that these reasons can be balanced against each other, and that in the majority of cases there is a fact of the matter about what an agent has most reason to do.²⁶

The Partial Dualist's suggestion that reasons sourced in the interests of others can be outweighed by considerations of self-interest is supported by common sense. And examples in which a minuscule self-sacrifice leads to an enormous increase in impartial welfare tell against Sidgwick's claim that rational adjudication in all cases of conflict between self-interest and the greater good is impossible. But the truth of Partial Dualism would imply that the axiom of Equal Good is false. If Parfit and Crisp are right, reason does not require giving equal fundamental weight to equal portions of the good, and my Sidgwickian argument for Rational Impartialism fails. Given the initial plausibility of Sidgwick's axiom, and given what's at stake for the Partiality Question, it will be necessary to further examine the merits of Equal Good, and also to delve deeper into the commitments and implications of Partial Dualism. This will allow us to determine whether Sidgwick's axiom and my argument for Rational Impartialism can withstand the Partial Dualist's challenge.

²⁵ Parfit believes the idea of 'normative imprecision' has been underappreciated by philosophers. He places substantial weight on this notion in his paper 'How We Can Avoid the Repugnant Conclusion' (unpublished).

²⁶ The view put forth by Samuel Scheffler in *The Rejection of Consequentialism* (1982) can be understood as a version of Partial Dualism. However, there is an important respect in which his view differs from the views of Parfit and Crisp. While Scheffler's 'agent-centered prerogative' is based on the fact that human beings typically care more about their own interests, Parfit and Crisp deny that reasons are grounded in facts about favoring attitudes. See Crisp (2006: 139 n. 38).

3 Explanatory Power and Criterial Arguments

One of the chief merits of the axiom of Equal Good is that it provides a compelling unifying explanation for the truth of a variety of normative judgments accepted by partialists and impartialists alike. Consider the following:

- (1) If forced to choose between a more painful experience tomorrow and a less painful experience today, you have decisive reason to choose the less painful experience today.
- (2) You have just as much reason to save the life of a person of a different race as you do to save a member of your own race.
- (3) If forced to choose between a nuclear bomb being dropped on either a large city or a small isolated village, you have decisive reason to choose the village.

In each of these examples, the choice deemed contrary to reason involves having unequal fundamental concern for equal portions of the good. For instance, in (1), the fact that one's hedonic welfare is no more or less valuable tomorrow than it is today, combined with the fact that one ought to have equal fundamental concern for equal portions of the good, explains why it is true that you would have decisive reason to choose the less painful experience today. The structure of this explanation is the same for claims (2) and (3).

In rejecting the axiom of Equal Good, Partial Dualists dispense with what appears to be a powerful unifying explanation for the truth of such judgments. This looks like a significant theoretical cost. The best way for Partial Dualists to overcome this worry is to offer an alternative principle that can also explain the judgments while remaining consistent with the central commitments of their view. In light of these considerations, and in light of the core

tenets of Partial Dualism, I suggest that Partial Dualists should advocate the following alternative principle:

Equal Good Relational: One ought to have equal fundamental concern for equal portions of the good, but only insofar as one is equally related to the different portions; different relations can justify different levels of fundamental concern.

The suggestion that different relations can justify having unequal fundamental concern for equal portions of the good is central to Partial Dualism. Recall that Parfit and Crisp's chief disagreement with Sidgwick concerns whether the existence of universal good must lead to an impartial account of practical reason. Though Partial Dualists accept that everyone's welfare is an equal part of universal good, they reject impartialism because they believe that the relations we have to ourselves are non-evaluative sources of reasons that compete with reasons sourced in universal good. In other words, although reason often requires that we give equal fundamental weight to equal portions of the good, differing relations can justify different assignments of weight.

Notice that Equal Good Relational is equally capable of explaining why the mere difference of a day is not reason to choose more pain rather than less (judgement 1 above). Since unequal fundamental concern for equal portions of the good can only be justified if one is differently related to those portions, and since there is no such difference in this instance, it would be contrary to reason to choose the more painful experience tomorrow rather than the less painful one today. Equal Good Relational is also capable of explaining the other judgments that seemed well explained by the unqualified version of Equal Good.

We now have two candidate principles for which we can make a pairwise comparison. If a stronger case can be made for Equal Good Relational, then my argument for Rational

Impartialism cannot succeed. But if it turns out that the original version of the axiom has greater justification, then the challenge from Partial Dualism is overcome.

A difficulty arises at this point, however. Given that Equal Good is a purported self-evident axiom of ethics, we cannot attempt to vindicate it by appealing to more basic principles in support of it. As Sidgwick points out in “The Establishment of Ethical First Principles,” it is not immediately obvious how philosophical progress can be made when there is disagreement over purported fundamental principles. He articulates the puzzle as follows:

On the one hand, it seems undeniable that first principles cannot stand in need of what is strictly to be called proof: they would obviously cease to be first principles if they were exhibited as dependent for their certainty on the acceptance by the mind of certain other truths. Yet, on the other hand, when we are dealing with any subject where there is a conflict of opinion as to first principles, we can hardly refuse to give reasons for taking our side in the conflict: as rational beings conversing with other rationals it seems absurd that we should not be able to explain to each other why we accept one first principle rather than another. (EEM 29)

Sidgwick goes on to outline two ways in which philosophical progress can be made when we encounter disagreement over first principles. The first involves addressing the principle of one’s interlocutor and trying to show that while there is some kernel of truth in it, there is some accidental or arbitrary feature that precludes it from being the true first principle (EEM 30). For example, if my interlocutor accepts that all pain of human beings is to be avoided, I can point out that the limitation to human beings is arbitrary, and so the real first principle is that all pain is to be avoided regardless of species membership (ibid.). David Phillips (2011: 63) labels these “bipartite ad hominem arguments” because they are addressed to particular interlocutors and they involve both a negative and a positive verdict on the interlocutor’s initial belief.

A second way of making progress regarding disagreement over first principles involves appealing to general criteria for assessing the relative strengths and weaknesses of the candidate principles. As Sidgwick explains: “We may be able to establish some general criteria for distinguishing true first principles (whether ethical or non-ethical) from false ones; and may then construct a strictly logical deduction by which, applying the general criteria to the special case of ethics, we establish the true first principles of this latter subject” (EEM 30). Following Phillips (2011: 64) I shall refer to arguments of this sort as “criterial arguments.”

The criteria that Sidgwick outlines in “The Establishment of Ethical First Principles” are essentially the same criteria he gives as part of his discussion of Philosophical Intuitionism in the *Methods*. As we saw in chapter 2, these criteria are as follows:

- (1) The terms of the proposition must be clear and precise.
- (2) The appearance of self-evidence must survive careful reflection.
- (3) The proposition must not conflict with other self-evident propositions.
- (4) The proposition must not be denied by an epistemic peer. (ME 338-41)

If Equal Good Relational appears no less evident than Equal Good, then neither passes Sidgwick’s third test. And neither principle appears to meet the fourth criterion, as evidenced by Parfit and Crisp’s rejection of Equal Good. Given Sidgwick’s belief that encountering such disagreement should reduce one to a state of neutrality, one might expect that no further progress can be made regarding the axioms. Recall, however, that Sidgwick qualifies his remarks about disagreement with the statement “if I have no more reason to suspect error in the other mind than in my own” (ME 342). Encountering disagreement over first principles does not mark the end of inquiry. If there are considerations one can raise which might explain why one’s interlocutor

rejects the candidate principle, it is worth articulating those considerations in hopes that the disagreement might be resolved. In the next two sections I shall argue that Equal Good fares much better than Equal Good Relational when exposed to Sidgwick's first two tests for highest certainty. If these arguments succeed, then our confidence in Equal Good need not be threatened by the disagreement.

4 Sidgwick's First Test

Sidgwick's first criterion for highest certainty requires that the terms of the proposition be clear and precise. As we saw in chapter 2, the axiom of Equal Good requires an initial reading of 'good' that is neutral between the universal and agent-relative conceptions. This seemed to indicate that Equal Good runs afoul of Sidgwick's first test because it lacks the requisite precision. But as I explained, the lack of specification in the axiom is not the type of problematic vagueness or ambiguity that the first criterion tests for. This is because neither specification threatens to undermine the plausibility or usefulness of the axiom by introducing arbitrariness or rendering it tautologous.²⁷ Moreover, the work done in chapter 3 (if successful) eliminates worries about ambiguity. Vindication of universal good means the axiom of Equal Good should be read with a universal specification of 'good'.

In order for Equal Good Relational to fair equally well by Sidgwick's first test it needs to be rendered more precise. Specifically, it needs to be made clear just what type of relations are such that they justify having unequal fundamental concern for equal portions of the good. Clearly, not just any relation will do. The fact I was born in the same hospital as another person does not justify having greater fundamental concern for that person than I do for people born in

²⁷ See pp. 49-50; 62 n. 40.

different hospitals. Nor does the fact another person and I share the same ethnicity provide reason to prioritize that person's interests over the interests of others. These sorts of relations are obviously normatively inert.

But just what relations do justify unequal fundamental concern for equal portions of the good? The answer is not immediately obvious. Hence, the application of Sidgwick's first test poses a challenge for the qualified version of Equal Good and for Partial Dualism itself. The challenge is to provide a specific answer to the question of what relations justify unequal fundamental concern, and to present a case for the ethical significance of these relations that is compelling enough to warrant deviating from the original version of the axiom.

4.1 Crisp on the Separateness of Persons

Crisp attempts to meet this challenge by appealing to the separateness of persons. The relation that Crisp believes justifies different levels of regard for equal portions of the good is the relation one has to oneself. In his chapter on practical reason, Crisp writes: "What grounds self-interested reasons is the value to the agent in terms of her own well-being (a value which could be understood as assessed from 'the objective point of view') *and* the fact that this value is instantiated in the life of the agent. That non-evaluative fact is part of the ground of the self-interested reason" (2006: 139).

In positing the separateness of persons as a non-evaluative source of reasons, Crisp does not appeal to the notion of rights or facts about how people ought to regard one another. Rather, he appeals to the descriptive fact that each individual possesses a distinct capacity for consciousness. The fact that we all have a separate capacity for conscious experience is what gives each of us "a special reason for promoting the enjoyment and minimizing the suffering

arising through that capacity's being exercised" (2006: 141). In an effort explain why an agent's separate consciousness matters in this way, Crisp quotes one of the passages from the *Methods* that we considered in previous chapters:

It would be contrary to common sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently 'I' am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the existence of other individuals. (ME 498)

Following this quotation, Crisp adds that "the difference between one person and another is, as Sidgwick says, real. However close I am to some other person, she will still be another person. And it is this difference which provides a source of self-interested reasons" (2006: 143).

While I do not wish to deny that there is a metaphysical difference between one person's capacity for consciousness and that of another, the problem is that it is unclear why this difference has any greater ethical significance than other differences between individuals such as differences of race, gender, nationality, etc. We can imagine a nationalist attempting to justify prioritizing the interests of his compatriots by saying "However close I am to a foreigner, she will still be a foreigner." Obviously the claim that we have ultimate reasons to favor ourselves has much greater prima facie plausibility than the claim that we have ultimate reasons to favor members of our own nation. But this is well explained by the fact that virtually all of us care more about ourselves than we do about strangers, and few of us experience the same disparity in caring between fellow countrymen and citizens of other nations. The passage from Sidgwick that Crisp offers for support is especially telling here. Given that I *am* more concerned about my own existence than about the interests of others, it is unsurprising that it seems as though I have "special reason" for promoting my own welfare. But for those with nationalist commitments it

will seem equally plausible that they have special reason for promoting the interests of their compatriots. Thus, it would appear that much more needs to be said about why my relation to myself has special significance that does not inhere in the various relations I have to other people.²⁸

These points are related to an objection Crisp himself raises against the appeal to social relations (e.g. friendships, familial relations, etc.) as a source of ultimate reasons. Crisp is skeptical that these relations are a source of ultimate reasons partly because they have a biological basis which conferred an obvious evolutionary advantage (2006: 143). The suggestion is that our natural inclination to favor our intimates makes intuitions about the rationality of such favoring less reliable.²⁹ But note that this worry applies equally well to self-favoring. Just as favoring the near and dear was adaptive for our ancestors, so was favoring oneself. Thus, if Crisp is right to be skeptical about social relations as a source of ultimate reasons, we should also be skeptical about one's relation to oneself as a source of ultimate reasons.³⁰

In addition to these epistemological concerns, Crisp's appeal to the separateness of persons also leads to worries about the general stability of his brand of Partial Dualism. Because he rejects the idea that kin relations and social relations are an independent source of reasons for an agent, Crisp must hold that there is something uniquely significant about one's relation to one's own consciousness. But the more significance we attribute to the separateness of persons, the more difficult it becomes to explain why pursuing one's own interests can be contrary to

²⁸ As we saw earlier, Crisp also appeals to intuitions about particular cases to support his view. I raise worries about this strategy in section 5 below.

²⁹ Crisp remarks that the evolutionary benefit of favoring kin is insufficient for debunking intuitions supporting ultimate reasons of partiality, but he claims (rightly, in my view) that it is enough to throw these intuitions into doubt (2006: 143).

³⁰ I return to this point in section 5 below.

reason (as it is when doing so comes at a great cost to others). Indeed, if my relation to myself is as radically different from my relations to others as Crisp suggests, it becomes difficult to see how any particular weighing of personal and impartial reasons could have greater justification than any other.

4.2 Parfit on Psychological Connections

Let us turn then to the specification of Equal Good Relational that is suggested by Parfit's discussion of partiality. Parfit's defense of Partial Dualism does not focus on each individual's metaphysically distinct capacity for consciousness, but rather the various psychological connections between one's present and future selves, as well the similar connections between oneself and one's loved ones (2011, 1: 136). Parfit believes these connections justify having greater regard for one's own good and the good of one's intimates than one has for the equal good of strangers. The relevant psychological connections are things like shared memories, desires, goals, intentions, etc.³¹

In order to understand why Parfit believes these connections justify unequal fundamental concern for equal portions of the good, we must consider a line of argument he presents in *Reasons and Persons* (1984). Parfit provides powerful arguments for the claim that when it comes to an individual's survival, it is not mere psychological continuity that matters but also the

³¹ David Brink (1997) also appeals to psychological connections in order to resolve Sidgwick's Dualism. According to Brink's 'metaphysical egoism', the psychological connections we have to others give us self-interested reasons to promote their interests. Since our own flourishing is dependent upon our relations to others as well as their flourishing, this shows that the ends of self-interest and the greater good are aligned after all. Though I don't have space to adequately address Brink's proposal, I can note here one major concern. The problem is that Brink's view does not explain how we can have reasons to promote the interests of strangers on the other side of the globe. As I argued in chapter 3, it's deeply implausible that the happiness of faraway people whom I shall never meet cannot provide *any* ultimate reasons for me. For further discussion of Brink's proposal see Singer and De Lazari-Radek (2014: 164-7).

persistence of the aforementioned psychological connections. He illustrates this point with the example of memory:

If our lives have been worth living, most of us value highly our ability to remember many of our past experiences. The loss of all these memories need not destroy continuity of memory, which only requires overlapping chains of memories. Suppose I know that, two days from now, my only experience-memories will be of experiences that I shall have tomorrow. On the view just stated, since there will be continuity of memory, this is all that matters. It should not matter to me that I shall soon have lost all of my memories of my past life. Most of us would strongly disagree. Losing all such memories would be something that we would deeply regret. (1984: 301)

Similar remarks apply for other psychological connections. Not only would I regret losing all of the memories of my past life, I would also regret the disappearance of my values, commitments, ambitions, etc.

Parfit makes a compelling case. It would indeed be very bad for an agent to lose all memories of her past even if she could maintain continuity through overlapping chains going forward. So it is plausible that these connections are crucial for meaningful survival and perhaps also for flourishing. But notice that this tells us nothing about why the individual should assign greater weight to her own survival and flourishing when deciding what to do than she does for the survival and flourishing of relevantly similar strangers. In *On What Matters*, Parfit seems to assume that because shared memories are necessary for meaningful survival they must also have significance for the question of partiality towards oneself. But this does not follow. The fact that my welfare partly depends on the persistence of my psychological connections does not imply that these connections provide reasons to prioritize my own welfare over the welfare of others.

Consider some of the other candidate psychological connections that might justify unequal fundamental concern for equal portions of the good. Could the fact that my future self shares some of the same desires and intentions as my current self ground reasons to prioritize my welfare over that of a relevantly similar stranger?³² The best reasons for skepticism about this possibility come from Parfit himself. In Part 1 of *On What Matters*, Parfit provides a forceful critique of desire-based theories of practical reason. He highlights the implausible implications of these views, and he explains why the standard strategies for mitigating the damage resulting from those implications are unsuccessful.³³ Parfit concludes that the mere fact that Φ -ing would satisfy an agent's desire (whether actual or idealized) is not a reason for that agent to Φ . But if desires themselves are normatively inert in this respect, why would the fact that I share desires with my future self have ultimate reason-giving force? If the fact that I desire to visit Paris is not by itself a reason to visit Paris, why is the fact that my future self will also have this desire a reason to prioritize my own welfare?³⁴

Similar remarks apply for shared intentions. Parfit denies that the mere fact that an agent intends to Φ is a reason for that agent to Φ . Whether an agent has reason to Φ , on Parfit's view, depends on the objective features of Φ -ing. It is only if Φ -ing is valuable for its own sake, or will lead to something of value, that an agent can have any reason to Φ . If this is right, why should the fact that an agent shares the intention to Φ with her future self be a reason for her to prioritize her own interests? If intentions themselves are normatively inert for an agent at a

³² By 'relevantly similar' I mean a stranger who is no more or less likely to benefit or harm others, and is likely to experience the same number of years of happy life as I am.

³³ Parfit (2011, 1: 58-110)

³⁴ In his unpublished manuscript 'Why Not Subjectivism?' David Sobel suggests that Subjectivism fits well with non-impartialist accounts of practical reason. The suggestion is that Subjectivism, combined with the fact that most of us care more about our own lives than the lives of strangers explains the widely shared intuition that we have at least as much reason to focus on our personal projects as we do to help needy strangers (p. 52). I think Sobel is right about the fit between Subjectivism and partiality, and I believe there is a parallel fit between Objective accounts of practical reason and impartialism.

particular time, it would be surprising if shared intentions have ultimate normative significance. Of course, there are cases in which it would seem that shared intentions do provide reasons. For instance, if my intention is to continue a promising line of medical research that might lead to a cure for cancer, then the fact that my future self will share this intention would be a reason to save my own life rather than the life of a stranger who intends to spend the rest of his life vacationing. But here the *source* of my reason is the valuable state of affairs I could bring about if my intentions are carried out successfully—it is not the mere fact that I am connected to my future self through the shared intention.

Recall from chapter 1 that philosophical investigation of the Partiality Question presupposes that our practical reasons are object-given and value-based.³⁵ It was Parfit's powerful critique of desire-based accounts that first helped me appreciate this point. The point of the previous two paragraphs has been to show that Parfit's commitment to object-given, value-based accounts of practical reasons is difficult to reconcile with his belief that psychological connections are a source of ultimate reasons. If psychological states such as desiring or intending are not sources of reasons for a particular individual, we should not expect that sharing these states with others (including our future selves) is a source of reasons.³⁶

³⁵ See pp. 19-27.

³⁶ In *The Point of View of the Universe* (Singer and De Lazari-Radek: 2014, 163) the authors cite an example Parfit provided to them in personal correspondence. The case involves stealing a stranger's raft in order to save one's own child, and thereby preventing the stranger from saving her own two children. Singer and De Lazari-Radek note that though Parfit initially claimed that such an act would not be contrary to reason, he has since indicated that he is reconsidering his view on this issue. This suggests that Parfit might be growing less confident about the extra rational weight provided my connections to oneself and one's intimates.

4.3 Summary

The Partial Dualist's qualified version of Equal Good does poorly when we apply Sidgwick's first criterion for highest certainty. When we attempt to give the principle the requisite precision by specifying the relations that justify unequal fundamental concern for equal good, serious problems emerge. First, it is not clear why a shared capacity for consciousness with oneself has greater ethical significance than any of the numerous other features we share with others. Second, the same reasons that warrant skepticism about intuitions supporting partiality towards kin also apply to intuitions supporting self-favoring (more on this point in the next section). Third, the idea that shared memories support ultimate reasons of partiality gains illicit support from the fact that connections to one's memories are important for personal identity. Finally, the purported significance of shared desires and intentions sits uncomfortably with Parfit's well motivated rejection of Subjectivism.

5 Sidgwick's Second Test

Recall that Sidgwick's second criterion for highest certainty requires that the intuitive plausibility of the principle survive careful reflection. Specifically, we need to investigate whether the intuitions supporting the principle might have originated from unreliable sources. As we have seen, Parfit and Crisp attempt to motivate their claim that we have ultimate reasons to prioritize our own good by appealing to common sense intuitions about cases of conflict between self and others. Presumably then, these are the sorts of intuitions that make it seem plausible that reason permits unequal weighing of equal portions of the good. The important question is whether these intuitions are to be relied upon. In the rest of this section I shall outline

four considerations which suggest that intuitions about cases of conflict do not justify abandoning Sidgwick's axiom in favor of Equal Good Relational.

5.1 Predominant Egoism

Psychological egoism is a descriptive view which holds that people can only be motivated to perform actions they believe to be in their self-interest. Although this view is implausible, there is a related view that is quite compelling. According to predominant egoism, genuine instances of altruism are rare, and they typically occur only when the sacrifice is small and the benefit to others is large, or when those benefited are loved ones or favorite causes.³⁷ Not only is predominant egoism supported by everyday experience and introspection, it is also well explained by evolutionary theory. In modern societies where individuals are protected by laws and other institutions, highly altruistic individuals are not at a significant reproductive disadvantage. But it is doubtful that the same was true in the early stages of human evolution. Given the ferocity of competition for resources and the lack of institutional safeguards, anyone who was not focused disproportionately on her own needs and those of her kin would be substantially less reproductively fit than individuals who had more self-centered attitudes and dispositions. It is highly plausible that the disposition to give considerable priority to self-interest was pervasive in the proto-human species we evolved from, and continued to be prevalent as our species evolved.³⁸

The plausibility of predominant egoism as a descriptive view about human motivations is highly relevant to the Partiality Question. This is because it is likely that the pervasive motivational tendency toward self-favoring has undue influence on our intuitions about reasons

³⁷ For discussion see Kavka (1986) and Shaver (1999).

³⁸ In addition to prioritizing self-interest, evolutionary pressures likely promoted prioritization of the interests of one's group over those of non-group members. See Sober and Wilson (1998).

for action. As beings that spend the majority of our time aiming to advance our own welfare it is unsurprising that we find it intuitive that self-interest carries at least some extra rational weight.³⁹ This is because intuitions about the rationality of conduct are often influenced by one's own dispositions, feelings, and motives.⁴⁰ Moreover, our judgments about what sort of conduct is reasonable are highly influenced by what is typical in the society we are part of. In *Reasons and Persons* Parfit points out that egoism has historically been the dominant view because the world's great religions teach that morality and self-interest ultimately coincide, and religious leaders have gladly appealed to self-interest to motivate people to obey the dictates of their religion. Parfit then suggests that this historical influence makes some of our intuitions less reliable: "Since S [Parfit's label for a theory which is essentially Rational Egoism] has been taught for more than two millennia we must expect to find some echo in our intuitions. S cannot be justified simply by an appeal to intuitions that its teaching may have produced" (1984: 130). Although Parfit's remarks are directed at intuitions supporting Rational Egoism, they apply equally well to self-favoring intuitions in general.

Here one might object that facts about our evolved motivational psychology can also cast doubt on intuitions supporting impartialist principles such as the axiom of Equal Good or Rational Impartialism. However, Peter Singer and Katarzyna de Lazari-Radek explain why this is not the case. Impartialist principles are immune from these worries because, as Folke Tresman (a noted proponent of evolutionary debunking arguments) acknowledged in correspondence with

³⁹ Of course, common sense intuitions also support our giving *some* ultimate consideration to complete strangers. The truth of predominant egoism is compatible with our assigning some ultimate weight to impartial interest because our normative intuitions are not determined *exclusively* by our inherited motivational tendencies. As the animals that can understand and respond to reasons, we are able to overcome our partialist impulses and recognize that things like suffering and achievement could matter even if they are occurring on the other side of the globe.

⁴⁰ Recall Sidgwick's claim that "it cannot be denied that any strong sentiment, however purely subjective, is apt to transform itself into the semblance of an intuition; and it requires careful contemplation to detect the illusion" (ME 339).

De Lazari-Radek and Singer, “we presently don’t have a fully satisfactory and well-established (evolutionary) debunking explanation of [these principles]” (2014: 194). While evolutionary theorists can explain our tendency to make sacrifices for others by appealing to kin altruism and group selection, there is no equally plausible explanation on offer for the judgement that we ought to give equal consideration to equal interests regardless of whether the interests at stake are one’s own, those of a stranger in a distant land, or even a member of a different species.

Guy Kahane (2011: 119) objects that impartialist principles are not immune from evolutionary debunking because they are merely a reasoned extension of principles arising from kin altruism and reciprocal altruism. The suggestion is that evolutionary pressures resulted in our aiming to help our kin and those who have helped us, and that in forming normative judgments we extended this altruism to include everyone so as to avoid incoherence. Since the starting point of these judgements (i.e. kin altruism and reciprocal altruism) is purportedly contaminated by natural selection, then the judgments themselves are unreliable. De Lazari-Radek and Singer responded by pointing out that impartialism can be reached by means other than a reasoned extension of partial altruism. As we have seen, for instance, Sidgwick’s principle of Rational Benevolence is deduced from the axiom of Equal Good (with a universal specification) and the axiom of Personal Irrelevance. This axiomatic deduction is an instantiation of a type of reasoning that is not merely an attempt to extend partial altruism to make it coherent.

A second objection from Kahane (2014) is that the appeal to evolutionary debunking arguments backfires for any theorist who wishes to maintain that pleasure is good. This is because if any normative belief is likely to be the direct result of selective pleasures, it is going to be the belief that pleasure is good. One way of responding to this worry is to note an

important difference concerning our epistemic situations with respect to the Partiality Question and the question of the value of pleasure. In the case of the former, there are strong intuitions on both sides. There is something intuitive about prioritizing one's own interests, and also something intuitive about giving equal weight to the equal interests of everyone. One way of making progress here is to consider which intuition is more likely the product of our ability to reason. The fact that partiality intuitions are explainable by evolution does not by itself debunk them. Rather, the combination of the evolutionary explanation and the existence of a strong competing intuition which is not equally explainable in evolutionary terms tell in favor of the competing intuition. Note that this is not the situation regarding the value of pleasure. We do not find ourselves with conflicting intuitions about the goodness of pleasure. Hence even if we should expect creatures with our evolutionary history to judge pleasure good, this does not give us sufficient reason to abandon this judgment because there is no competing intuition supporting the claim the pleasure is not good.

5.2 Derivative Reasons in Practice

A second source of skepticism about partiality intuitions comes from the fact that, regardless of which theory of rational weight is correct, in practice almost all of us have reasons to focus disproportionately on ourselves and our loved ones. Given that each individual is in a unique epistemic and practical position with respect to her own needs, it would be highly inefficient for individuals not to pay special attention to their own basic needs before turning to the needs of others (in normal circumstances). Although it is just as important for my colleague to eat a nutritious breakfast as it is for me, it would be a waste of resources to drive to his house in the

morning to cook for him. For the sake of efficiency we should each focus on our own nutritional needs first. The same is true for numerous other aspects of our lives.⁴¹

Similar remarks apply in the case of friends and family. Even if the needs of a stranger on the other side of the globe carry no greater *ultimate* rational weight for me than the needs of my friend, I am obviously in a much better practical and epistemic position to meet the needs of my friend. Moreover, there are certain needs my friend has that only someone who knows her well can provide (e.g. emotional support). Furthermore, given the facts about human psychology, it is plausible that for most of us, failing to maintain intimate relationships in which we pay especially close attention to the needs of certain individuals would be detrimental to our mental health and well-being. And it is plausible that jeopardizing our own health and happiness in this way would be suboptimal from a purely impartial perspective. It is much easier to promote the greater good of others when we are healthy, happy, and productive.⁴² Thus, in practice, there is a significant disparity between the rational weight of the interests of my friend and the interests of a stranger in a faraway land. But it does not follow from this fact that the interests of my friend are weightier for me at the fundamental level.

In light of the obvious fact that individuals should focus on their own basic needs first, and that we are often in a much better position to help those close to us, it is easy to see how we might find it intuitive that we are rationally justified in displaying fundamental partiality. But to the extent that partiality intuitions are influenced by the fact that in practice we should focus disproportionately on ourselves, these intuitions are dubious. This is because considerations of how we ought to focus our efforts in practice are not relevant to the theoretical question of ultimate (i.e. non-derivative) rational weight. The fact that individuals should focus a

⁴¹ Sidgwick notes this point in discussing utilitarianism and his principle of Rational Benevolence (ME 252, 382).

⁴² See Railton (1984).

disproportionate amount of attention on their own basic needs can be accommodated just as well by impartialist theories as it can by Partial Dualism.

5.3 Different Senses of 'Irrational'

Debates about the rational conflict between self-interest and the greater good often employ the term 'irrational'. We ask, for instance, whether it would be irrational for an agent to buy herself some luxury item rather than donating that money to famine relief. I believe one of the reasons why some level of self-favoring strikes many as intuitive is the slipperiness of the term 'irrational'. This word is widely used in several different senses, and it is crucial to recognize which sense is most important for the Partiality Question.

When people describe an agent or action as irrational, they often mean to say that the agent or action is 'senseless', 'stupid', 'idiotic', or 'crazy' (Parfit: 2011, 1: 33). Following Parfit, we can call this the non-technical sense of 'irrational'. At other times, people use the term 'irrational' to describe an agent who fails to take the necessary means to her own ends. We can call this the instrumental sense of 'irrational'. A third sense of 'irrational' is the subjective sense. An agent is said to be irrational in this sense when she fails to do what she believes she has most reason to do. This is related to the final sense of 'irrational' which is the objective sense. An agent is irrational in the objective sense when she fails to do what she in fact has most reason to do.

For the purpose of investigating the Partiality Question, the most important sense of 'irrational' is the objective sense. The aim of our inquiry is to discover facts about our objective reasons. We are not interested in facts about whether someone is crazy, or failing to take the means to her own ends, or failing to do what she believes she has most reason to do. Yet when

we consider cases of conflict between self-interest and the greater good, it is easy to mistakenly focus on one of the other senses of ‘irrational’. Of course we should find it obviously false that an agent who chooses to save her own life rather than the lives of two strangers is thereby irrational, if by this we mean that she is a fool, or that she has failed to take the means to her own ends. But the relevant question is whether by saving herself she failed to do what she had most objective reason to do. The fact that we often mistakenly use the term ‘irrational’ in one of these other senses should make us less confident about intuitions supporting the partialist answers to hypothetical conflict cases. One might object to this point on the grounds that nobody is as keenly aware of the significance of the different senses of ‘irrational’ as Parfit and Crisp. However, I believe the preceding discussion does go some way towards undermining self-favoring intuitions for two reasons. First, even those who are well aware of important conceptual distinctions are not immune from having their intuitive judgments unduly influenced by the wrong concepts. Second, one of the theoretical advantages of Partial Dualism is the fact that the intuitions supporting self-favoring are widely held. Even if the intuitions of chief proponents of the view are not unduly influenced by less important senses of ‘irrational’, the fact that the judgments of so many others probably are distorted by the wrong concepts is relevant.⁴³

5.4 Taking ‘Morality’ too Seriously

Philosophers have long viewed adjudication of conflicts between partial and impartial interest as a matter of determining what morality requires of us. As I explained in chapter 1, I believe this

⁴³ There is some reason for suspecting that even Sidgwick may have had his intuitions distorted by the fact that egoism is compatible with being fully ‘rational’ or ‘reasonable’ in several senses. Hastings Rashdall suggests that something along these lines occurred in Sidgwick’s writing of the *Methods*. When discussing Sidgwick’s claim that self-interest is a reasonable object of pursuit even when it conflicts with the general good, Rashdall writes, “[Sidgwick] seems to be relapsing into that meaning of the term ‘reasonable’ which has generally found favor with Hedonists who do not profess to be ‘rationalistic’ –that is to say, ‘internally self-consistent’ or ‘conducive as a means to the end which any one happens actually to desire’” (1907: 57).

tendency to focus on moral obligation when thinking about questions of rational weight has had a distorting influence on our intuitions. To briefly recapitulate, the concept of moral obligation (which is central to moral discourse) is closely tied to the notion of blameworthiness. To say that an action is morally wrong, on one common conception, is to say that an agent who performs it is blameworthy, which amounts to her meriting emotional sanctions in the form of resentment from others and feelings of guilt within herself.⁴⁴

Insofar as we have a tendency to think about the conflicts between partial and impartial interests in terms of moral permissibility, it is unsurprising that intuitions often support partiality. In light of facts about human nature, such as our strong instinctual drive to preserve our own lives, it seems extreme to suggest that an individual who chooses to save her own life rather than the lives of two strangers has done something which merits resentment from others and feelings of guilt within herself.⁴⁵ This is one of the major reasons why completely impartialist moral theories are often regarded as implausibly demanding. Many people find it intuitive that a minimal degree of partiality could not be blameworthy. If it is, then we get the implausible result that virtually all of us are, as Thomas Nagel puts it, “miserable sinners” (1986: 202).

I believe this fact about our moral intuitions partly explains the tendency for intuitions to tell in favor of partiality when considering cases of conflict. While many would deny that moral requirements are rational requirements, it seems like a reasonable conjecture that most people assume that if an action or attitude is morally permissible, it is thereby rationally permissible.⁴⁶

⁴⁴ See references in Ch. 1, note 13.

⁴⁵ Of course, many moral theorists believe that there can be blameless wrongdoing. Nonetheless, when we make intuitive judgments about whether an action would be wrong, it is likely that considerations of blameworthiness typically play a role, given the ubiquity of the blameworthy/‘reactive attitude’ conception of moral wrongness.

⁴⁶ This idea plays a crucial role in Dale Dorsey’s recent work on the question of the rational authority of morality (2012, ms).

Since we judge that partiality is morally permissible, we are disposed to judge that it is rationally permissible as well.

The fact that intuitions about the greater rational weight of partial interests are likely influenced by our judgments about the moral permissibility of partiality should diminish our confidence in them. The Partiality Question must be kept distinct from questions of moral obligation. When we ask whether self-interest is a source of ultimate reasons with greater weight than that of reasons sourced in the welfare of others, our concern is not about the moral status of individuals or their actions. The question is not whether an agent who gives fundamental priority to herself or her family is blameworthy—it is whether she failed to do what she had most reason to do.⁴⁷

5.5 Summary

I have offered four distinct grounds for skepticism about the reliability of intuitions supporting fundamental reasons of partiality: (1) the plausibility of predominant egoism, (2) the undue influence of facts about derivative reasons, (3) conflation of different senses of ‘irrational’, and (4) a misguided emphasis on morality when considering cases of conflict. I do not claim that these considerations alone undermine Equal Good Relational or Partial Dualism itself. My aim has merely been to show that intuitions supporting the partialist option in cases of conflict provide unstable grounds for deviating from Sidgwick’s axiom. Were it not for the existence of an apparently self-evident principle suggesting that the good is to be aimed at impartially, our self-favoring intuitions might be enough to justify Partial Dualism in spite of the doubts I have raised. But the axiom of Equal Good, being immediately cognizable by abstract intuition and

⁴⁷While such failure can merit criticism, this criticism ultimately amounts to informing the agent that she has acted suboptimally. This is different from moral criticism because there is no conceptual link between suboptimal performance and meriting feelings of guilt or shame. Cf. Brian McElwee (2007).

capable of withstanding careful reflection, raises the standards of justification for those who wish to vindicate ultimate reasons of partiality towards oneself. Intuitive judgments about cases of conflict, which are problematic for the reasons I have outlined, are inadequate for this task.

7 Conclusion

Despite their belief in agent-neutral good, Partial Dualists resist impartialism because they deny that one ought to have equal fundamental concern for equal portions of the good. I have tried to show that their rejection of Sidgwick's axiom stands in tension with their own commitments, and that the intuitive judgments that constitute the primary support for Partial Dualism are dubious. I conclude that if proponents of Partial Dualism wish to maintain their commitment to agent-neutral good and an objective theory of practical reason, they should embrace the axiom of Equal Good and the impartialism to which it leads.

I have now defended each of the controversial premises of my Sidgwickian argument for Rational Impartialism. However, the task of vindicating RI is not yet complete. Despite the arguments presented in the preceding chapters, I suspect that many will insist that Rational Impartialism has implications that are too paradoxical to accept. The aim of the next and final chapter is to remove the air of paradox surrounding RI.

5 Derivative Reasons and the Air of Paradox

1 Introduction

The conclusion of my Sidgwickian argument states that one ought to have equal fundamental concern for the equal good of all individuals. Though I hope to have made a convincing case for the truth of the premises leading to this conclusion, there is more work to be done. During the past three years I have discussed Rational Impartialism with numerous individuals—both philosophers and laypersons alike. Nearly everyone who I have informed of my assent to RI has responded in an unsympathetic fashion.¹ The responses have ranged from puzzlement and skepticism to incredulity and even mockery. Philosophers typically respond by immediately proposing ostensible counterexamples to the view: “So you’re saying it’s irrational for me to be at this conference right now?” Others have responded by questioning whether a life lived in accordance with RI could be a life worth living. Still others have wondered how my belief in RI can be reconciled with my choosing to pursue a Ph.D. in philosophy rather than doing something that would presumably bring about much more total good, such as becoming an investment banker and donating my income to charity.

Given these reactions, I worry that the seemingly paradoxical implications of RI might be so troubling that even readers who find the considerations raised thus far persuasive will still be convinced that something must be amiss. It is for this reason that I shall devote this final chapter to diminishing the air of paradox surrounding RI.

I begin in section 2 by considering the objection that Rational Impartialism is implausibly demanding. I consider some of the most challenging cases for RI, and I explain how our

¹ One of the notable exceptions was Neil Sinhababu, who responded by giving me a hug.

tendency to ignore important facts about such cases can have a powerful distorting effect on our judgements. In section 3 I consider the standard objection that impartialist views alienate us from the things we care about most and that adhering to such a view would prevent us from living meaningful lives. I explain how such worries are mitigated by emphasizing the role of derivative reasons as well as the distinction between motivating reasons and normative reasons. I also stress the point that we should not have expected the true account of practical reasons to happily coincide with our individual desires, goals, and ambitions in the first place. Finally, in section 4, I respond to the worry that RI cannot accommodate common judgements about cases of altruism in which an agent sacrifices her own greater good for the lesser good of another.

2 Demandingness

2.1 The Demandingness Objection

Worries about demandingness are traditionally raised against versions of consequentialism such as utilitarianism. Critics claim that by positing maximization of general happiness as the threshold for moral permissibility, utilitarianism is an implausibly demanding view. Many theorists find it hard to accept that we could be acting impermissibly each and every time we fail to act in a hedonically and impartially optimific manner. Since virtually no one comes close to living up to this standard, utilitarianism seems to imply that even the best among us are almost always acting immorally. And since we typically think of immoral or impermissible conduct as blameworthy, this means that if utilitarianism is true, we are all deserving of substantial amounts of blame, and even ostensible moral exemplars ought to feel guilty for many of their choices insofar as they fall short of maximizing universal happiness.

One might expect that the exact same objection can be leveled against proponents of Rational Impartialism. Recall, however, that RI is not a theory of morality. The view is not committed to the claim that whenever an agent shows greater fundamental regard for her own welfare or the welfare of her intimates she has therefore done something morally wrong, and hence that her conduct merits feelings of anger from others and guilt or shame within herself. All Rational Impartialism implies about such an agent is that she has failed to respond appropriately to her practical reasons. Hence, insofar as the demandingness objection concerns the issue of implausibly high standards for avoiding blameworthiness, the objection has no force against RI.

That being said, there is a version of the demandingness objection that can be fairly raised against RI. One might argue that theories of ultimate practical reasons can be implausibly demanding in a different way. Even though RI does not imply that an agent is blameworthy for prioritizing her own good, the view does imply that such an agent exhibits a type of irrationality in the sense that she fails to respond appropriately to her reasons. And given that most of us seem to give at least some fundamental priority to our own good, this would suggest that almost all of us are acting irrationally most of the time. Critics might charge that an account requiring fundamental impartiality in order to avoid acting irrationally demands too much of agents for it to be taken seriously—it is just not plausible that irrationality is as pervasive as RI implies. The demandingness objection can also be pressed against RI by highlighting specific examples of conflict between personal interests and the interests of strangers for which we find it implausible that an agent would have decisive reason to choose the impartially optimistic option due to the high level of sacrifice required. In the rest of this section I shall address both of these manifestations of the demandingness objection.

2.2 Pervasive Irrationality

Let us first consider the suggestion that RI is unlikely to be true because it is implausible that irrationality is as pervasive as the view implies. As I noted in chapter four, the term ‘irrational’ is used in a variety of senses. The sense that is relevant here is what I have called the ‘objective’ sense. Given that most of us fail to give equal weight to the equal good of everyone, RI implies that we are quite often behaving irrationally in the objective sense. In other words, most of us fail to do what we have most reason to do most of the time. Some might take this implication of RI as grounds for skepticism about the view because we tend to think that, as the animals that are capable of understanding and responding to reasons, it is unlikely that we could be failing so miserably.

There are two lines of response that I believe undermine the force of this objection. First, though I agree that very few of us are living an optimally rational life by the lights of RI, the extent to which we fall short of this is not as great as it might appear at first blush. As I explained in the previous chapter, regardless of which theory of practical reason is true, in practice almost all of us have reasons to focus disproportionately on ourselves and our loved ones. This is explained by epistemic and practical considerations, as well as facts about human psychological and emotional needs. Not only am I uniquely situated to promote the needs of myself and those around me, by doing so I make it easier for my intimates and myself to effectively promote the greater needs of others.² Of course, even taking these derivative reasons into account, most of us still fall well short of living in accordance with the reasons posited by RI. Nonetheless, we are not as far off the mark as one might have thought initially.

² See Railton (1984).

The second response to the worry about pervasive irrationality is to note that, setting the Partiality Question aside for the moment, most everyone accepts that they fail to respond appropriately to their practical reasons with high regularity. This can be seen by considering some uncontroversial examples of the things that most people do not do despite having strong reasons to do them. For instance, the vast majority of us have strong reasons to change at least some aspects of our diets, as well as reasons to exercise more. Yet despite our annual New Year's resolutions to exercise regularly and cut back on junk food, the majority of us fall back into our unhealthy habits by June if not sooner.³

To take another example, unless egoism is true, it is clear that we have strong reasons to donate a substantial amount of our disposable income to reputable and effective charities. Impartialists, Partial Dualists, and common-sense moralists can all agree that the great suffering occurring around the globe provides strong reasons to forego many of our luxuries in order to make a financial contribution. And most of us would admit that we fail to donate as much as we should.⁴ Given this, on any of the widely accepted theories of practical reason, most of us act irrationally in the objective-reason sense with great regularity.⁵ Of course, if RI is true, we are falling shorter of the mark than we are if a view like Partial Dualism is true. But since the rational failure is pervasive in either case, the rational failure implied by RI not a compelling reason for skepticism.

³ Lenny Bernstein, "It's a Week into January and a Quarter of Us Have Already Abandoned Our New Year's Resolutions," *The Washington Post* (January 2015): <http://www.washingtonpost.com/news/to-your-health/wp/2015/01/07/its-january-7-and-a-quarter-of-us-have-already-abandoned-our-new-years-resolutions/>

⁴ Although Crisp believes everyone has ultimate reasons to prioritize their own welfare, he acknowledges that given the present state of the world (i.e. given the existence of extreme global poverty), "the lives of most of us are unjustifiable in the light of reason" (2006: 141).

⁵ Even if egoism is true, practical irrationality is pervasive insofar as we consistently engage in imprudent behavior due either to weakness of will or to a mistaken belief that we ought to sacrifice our welfare for the sake of others.

A third example will further drive home the point. One of the counterintuitive features of RI is the implication that personal relationships are only of derivative importance. Common sense suggests that close personal bonds are highly valuable for their own sake. Perhaps no relationship is held in higher regard than that between mothers and their children. For most of us, once we reach adulthood and move out into the world on our own, we develop a strong sense of gratitude towards those who raised us—especially our mothers. And most mothers of adult children derive great enjoyment from visiting with them or at least talking on the phone. This explains the common belief that one ought to call or visit one’s mother with high regularity. But despite these normative beliefs, most of us fail to call or visit as much as we judge that we should. Given that common sense suggests that one ought to call one’s mother regularly, the fact that many of us do not further illustrates the point that even on a common-sense account of practical reasons, irrationality in the objective sense is pervasive. Hence, the pervasive irrationality version of the demandingness objection is not a serious threat to RI.

2.3 Hard Cases

A second way of pressing the demandingness objection against RI is to point to specific cases in which the view requires an implausibly high level of self-sacrifice or impartiality in order to act in accordance with one’s reasons. Consider, for instance, the following case from Doug Portmore.⁶ Suppose you have happily been working in the same corner office in the same building for the past several years. And suppose that a new faculty member has been hired, and she informs you that it would make her very happy if you would give her your office and move into the smaller, windowless office that was initially offered to her. If RI is true, then you ought

⁶ Portmore presented a version of this case during personal communication at the 2014 Rocky Mountain Ethics Congress. It is similar in structure to some cases he presents in *Commonsense Consequentialism* (2011: 4, 26, 30-31).

to give equal weight to the equal good of everyone. Hence, if it is the case that giving up your office would lead to even the slightest net increase in total good, an increase of say, one *utile*, then RI implies that this is what you have most reason to do.⁷ Portmore finds this implication of RI absurd. The thought is that the mere fact that giving up your office would result in a minuscule increase in impartial good cannot be enough to give you decisive reason to do so.

Though I understand why the above implication of RI might initially appear absurd, I believe this appearance diminishes when we consider what would need to be true in order for it to be the case that swapping offices would lead to an increase in overall good. It is tempting to think of the example as being based on the premise that the new faculty member would get slightly more enjoyment from your office than you do. But this would not be sufficient for it to be true that trading offices would lead to a net increase in total good. This is because packing up all of your belongings and moving them into a different office requires a great deal of time and energy. Moving offices is a substantial inconvenience, and it is typically an unpleasant experience. Moreover, your future working hours will presumably be less enjoyable for you in the windowless office, given that you had grown used to the much nicer corner office. Thus, given the amount of self-sacrifice involved in making the exchange, it would have to be the case that your new colleague would experience a very substantial increase in welfare from moving into the corner office in order for the trade to be overall good. This might be possible if, for instance, moving into your office would somehow help to alleviate an episode of depression that would otherwise prevent your colleague from experiencing happiness for the next several weeks. But if it is circumstances such as these that would allow for the exchange to be overall good, it

⁷ This is not to say that RI implies that one ought always to maximize total happiness by any available means. RI is officially compatible with non-welfare based reasons including deontological constraints. To simplify things we can suppose that in the above example, the only normatively relevant factors in play are the welfare of you and your colleague.

no longer seems absurd to suggest that you would have decisive reason to move offices. On the contrary, given that your colleague's happiness matters every bit as much as your own, and that the cost to you (taking everything into account) is less substantial than the benefit to her, it seems clear that you would have decisive reason to make the trade.

A second difficult case for Rational Impartialism comes from Crisp in *Reasons & the Good* (2006). In order to show that there is a limit to the amount of self-sacrifice one could have reason to make for the sake of others, Crisp provides the following example:

Two Doors 7. You are currently living a reasonably good life. You are now confronted by two doors. If you pass through door A, your global well-being will drop dramatically, so that the rest of your life will be worth living but of a very low quality. If you pass through door B, nothing will happen to you, but the future well-being of some other person, a stranger and out of sight, also living a reasonably good life, will drop even further, so that her life is barely worth living. (2006: 140-41)

If RI is true, then you have decisive reason to pass through door A. Of course, if you found yourself in such a situation you would likely believe that you have a reason to pass through door B. In discussing a case involving a similar conflict between self-interest and the good of others, Crisp (2015: 198) acknowledges that it is available for proponents of RI to appeal to evolutionary psychology to explain why we tend to judge that we would have a reason to pursue the self-interest option in cases such as this. Recall that I employed this strategy in discussing Partial Dualism in the previous chapter.⁸ However, Crisp suggest that a plausible explanation for why you would think you have such a reason is available without appealing to evolution—the explanation is that you would in fact have such a reason (ibid. 199). This is taken to be strong grounds for skepticism about RI.

⁸ See pp.109-112. For a comprehensive attempt to utilize evolutionary considerations to support impartialism and utilitarianism see Singer and de Lazari-Radek (2014: ch. 7). For a critique of their strategy see Crisp (2012).

Rather than rehearse the reasons for skepticism about self-favoring intuitions, I think it would help to consider a variation of *Two Doors 7* that eliminates some of the distorting influences of our motivational psychology. In chapter 3 we considered an example from Simon Keller involving a unique civilization that functions quite well despite the absence of close personal relationships. This was instructive for diminishing intuitive support for the claim that close relationships are valuable for their own sake. I believe that a similar idea can be useful for undermining the force of Crisp's thought experiment.

Imagine that there exists, on a faraway planet called Rashdall, a species of intelligent beings who are similar to human beings in most significant respects. The only major difference is that, as a result of a different evolutionary history, the Rashdallians do not have the innate drive towards self-favoring and kin partiality that are characteristic of human beings. When individual Rashdallians face conflicts between self-interest and the greater good, they typically do not experience an inclination to choose the former. While Rashdallians have the same sense of self that human beings typically possess, and while it matters to them that their own lives go well, their innate motivations are not biased toward themselves and their loved ones. But neither are Rashdallians strongly inclined to choose the sacrificial option. Rashdallians have the peculiar feature of possessing no innate motivational tendencies when they encounter such conflicts. At every such conflict they simply reflect on the balance of reasons in favor of each option. To avoid begging any questions, we will not assume which option is favored by the balance of reasons.

Now imagine that there is a Rashdallian who encounters a situation similar to *Two Doors 7*. Everything is the same as in the original version of the case except that it is now Rashdallians rather than human beings who are involved. The crucial question remains the same: Does the

Rashdallian have most reason to pass through door A, thus causing her own well-being to drop to a very low quality, in order to prevent an even worse outcome for a stranger? Or does she have most reason to pass through door B?

Before answering this question it is worth explaining why we should bother considering an imagined case involving alien beings. The reason the Rashdallian version of the example is significant is that it allows us to overcome some of the distorting influences that arise when we consider thought experiments involving human beings. These distorting effects take two forms. First, because we are human beings, it is difficult to prevent our own motivational and emotive responses from influencing our judgments about thought experiments involving drastic outcomes for other humans. When we are asked to judge what the agent facing the imagined set of circumstances has most reason to do, the impulses that might unreflectively pull us in one direction or another were we in those circumstances can also influence our second-person assessments. This effect is made even worse when we are asked what we ourselves would do if faced with such a decision, as in Crisp's *Two Doors* 7. When I imagine that I am the one who faces the choice between giving up my current happy life to drop to a level of well-being that makes my life barely worth living, I experience slight feelings of anxiety and dread (even as I sit comfortably in the air-conditioned library). It is only to be expected that my powerful instinctual aversion to such outcomes will have at least some influence on my intuitions about which option is supported by weightier reasons.

A second distorting effect that arises when we focus on examples involving human beings is the influence of one of the less important senses of 'rational' which I have discussed in previous chapters. We know that human beings are highly self-interested. Though many of us are inclined to take an interest in the well-being of others, few of us make the promotion of the

general welfare our guiding principle. Given that most of us are primarily focused on ourselves and our intimates, we assume that an act of great sacrifice would frustrate an individual's chosen ends. This makes such sacrifice an instance of instrumental irrationality. The fact that we tend to assume that an individual who chooses door B would be instrumentally irrational insofar as she would be frustrating her primary end of promoting her own welfare can plausibly bias our intuitions in the direction of door A.

By focusing on Rashdallians we can avoid both of these distorting effects. Since it is made salient that Rashdallians are different from us in that they do not have an innate disposition to prioritize their own interests, it is easier to focus solely on the issue of which option the agent has more reason to choose. It should be uncontroversial that the Rashdallian has reason to care about her own well-being. And only proponents of egoism would deny that she had reasons to care about the well-being of the other Rashdallian whose welfare hangs in the balance. The question then is what the relative weight of these reasons are (i.e. the Partiality Question).

When I reflect on this alternative version of the thought experiment it is evident to me that the Rashdallian has decisive reason to pass through door B. Although this choice will make her own situation extremely bleak, if she does otherwise another Rashdallian will suffer an even worse fate. And the suffering of this other individual will not be any less bad simply because it is happening to someone else. The fact that the Rashdallian does not have any innate tendency to prioritize her own welfare makes it easier to focus on what really matters in the case which is the welfare at stake. By thinking about what a Rashdallian would have reason to do rather than what a human being would have reason to do, or worse yet what "I" would have reason to do, we can eliminate the distorting influence of our desires and inclinations. And given the assumption that subjectivism is false, the judgements we endorse about the reasons had by alien beings with

different motivational psychologies should not differ from the judgements we endorse concerning the reasons possessed by human beings facing similar circumstances. If we agree that the Rashdallian would have most reason to pass through door A, then we should say the same thing about a human being who faces such a choice.

3 Alienation

The demandingness objection is not the only traditional critique of utilitarianism that can also be raised against Rational Impartialism. Critics of RI might also appeal to worries about alienation, which is a second standard line of objection against utilitarianism. The basic idea is that adhering to an impartialist ethical theory would alienate us from our true selves and the things that are most important to us, such as our loved ones and our personal projects. Hence, unlike the demandingness objection, the alienation worry takes a similar form against both utilitarianism and RI. This is because concerns about alienation directly address the impartialism that is inherent in both theories.

Worries about alienation can be expressed in various ways, but for present purposes it will suffice to focus on two particular versions of the objection. The first holds that impartialist views preclude individuals from having integrity and living a worthwhile life. The second is that impartialist views are incompatible with appropriately responding to the value of the individuals around us. I shall respond to each of these worries in turn.

3.1 Integrity

The integrity objection is most often associated with Bernard Williams. Though Williams presses the objection against utilitarianism, many of his claims could be applied just as well to Rational Impartialism. Williams' primary complaint is that any view requiring impartiality at the

fundamental level implies that a person's life is not really his or her own. He focuses specifically on the importance of choosing one's own projects and commitments, and the fact that impartial ethical views require that agents be willing to set aside their personal projects and convictions whenever doing so would be impartially best. Williams finds this aspect of impartialism absurd. In an oft-quoted passage, he puts the point thus:

The point is that [the agent] is identified with his actions as flowing from projects or attitudes which...he takes seriously at the deepest level, as what his life is about...It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimistic decision; but this is to neglect the extent to which *his* projects and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity. (1973: 116-17)

These remarks can be interpreted in at least two different ways. On one reading, Williams' objection is that it is absurd to suggest that an agent could have reason to sacrifice the very things that he most strongly identifies with because it would be impossible to be motivated to make such a sacrifice. Understood this way, the objection is easily met by appealing to the distinction between motivating reasons and normative reasons.⁹ It might be true that some agents could never be motivated to sacrifice their most sacred life projects for the sake of distant strangers. But that is consistent with the agent's having decisive normative reason to make the sacrifice nonetheless. Normative reasons are considerations that count, for the agent, in favor, of performing an action.¹⁰ And if we continue to work under the assumption that subjectivism is

⁹ Singer and de Lazari-Radek (2014: 135).

¹⁰ Crisp (2006: 38). The normative reasons we are here concerned with are, of course, practical reasons rather than epistemic reasons.

false, and that normative reasons are object-given, then we can accept that an agent can have a reason to Φ even if Φ -ing is not in any way connected to her subjective concerns.

On a second interpretation, Williams' objection is a first-order claim about the reasons an agent might have, rather than a meta-ethical claim about the nature and sources of one's reasons. On this alternative reading, Williams' claim is that a willingness to sacrifice one's personal projects whenever doing so is impartially optimal is incompatible with maintaining one's integrity. If we are required to set aside the things we care most about whenever their maintenance requires a certain sacrifice of the greater good of others, then we cannot live authentic and meaningful lives.

The problem with this objection is that Williams seems to be championing a subjective notion of integrity according to which one ought to be committed to carrying out one's projects, whatever they happen to be. But it is highly implausible that one has reasons to carry out whatever personal projects one happens to have adopted. If having integrity simply means staying committed to one's principles and projects regardless of their effects on others, then it is far from clear why we should think there is anything counting in favor of integrity. As Elizabeth Ashford (2000: 423-4) argues, if we appeal to this subjective notion of integrity we quickly run into absurd implications. If I happen to have structured my life and my identity around wanton cruelty, then maintaining my integrity would necessitate my continuing to cause suffering to innocent individuals. But this would be no reason to do so.

One might try reviving Williams' objection by arguing that there is special value inhering in staying committed to one's projects whenever these projects are objectively worthwhile. There is certainly something plausible about this suggestion. It does seem important for a person's wellbeing that she can immerse herself in the things that matter to her,

and develop a sense of identity from the worthwhile projects she commits herself to. But if this is right, the value inhering in integrity is sourced in considerations of welfare. And nobody's welfare is any more valuable than anyone else's. So even if staying committed to my projects is good for me, that does not imply that I have reason to maintain this commitment if doing so leads to a sacrifice of the greater good of someone else who is no less deserving of happiness than I am.

Here one might object that maintaining commitment to one's projects is not just one factor among many that constitute a person's welfare. Perhaps it is the case that such commitment is a necessary condition for a person's life to be worth living at all. There are two replies to this worry. First, it is highly implausible that a precondition for a person's life being worth living is that she view her projects as something she would not give up for the sake of the greater good of others. The sense of meaning I derive from my personal projects does not depend on my viewing them as non-negotiable. The projects and commitments that shape my identity and structure my day-to-day life provide meaning for me in virtue of their intrinsic features and the enjoyment and sense of accomplishment they give rise to. If it were the case that I was constantly changing my commitments and projects, or if I was under constant threat of losing the mental and physical capacities that allow me to engage in them, that would certainly detract from my sense of contentment and meaningfulness. But awareness of the fact that I might have decisive reason to abandon these projects or commitments for the greater good of others who are no less important than I am does not detract from the sense of meaning provided by my projects.

Second, we must bear in mind that Rational Impartialism is in principle compatible with agents pursuing their own personal projects. Pursuit of one's projects is among the chief means

of pleasure and achievement, which are two of the most plausible constituents of a person's good. But we must be careful not to confuse the question of what an agent has most reason to do, with the question of what would be the best possible life for a human being. It is certainly plausible that the best possible human life would include plenty of time and resources allocated to personal projects and relationships. In an ideal world, everyone would have the resources necessary to immerse themselves in meaningful projects, whether in the arts, science, athletics, or any of the other endeavors that would bring enjoyment and enrichment. Unfortunately, we do not live in an ideal world. Given the level of suffering that is taking place around the globe, reason requires most of us to make substantial sacrifices of our own good. This is because no individual's good is any more valuable than anyone else's, and as rational beings we are bound to have equal fundamental concern for equal portions of the good.

3.2 The Value of Individuals

A second version of the alienation worry holds that impartialist views alienate us from those around us insofar as adhering to such views precludes us from responding properly to the value of individuals. A clear articulation of this view is found in the work of David Velleman. On Velleman's view, although every rational agent is of equal intrinsic value, if you respond to everyone's value in a like manner, you fail to respond appropriately to anyone's value. The value that inheres in individuals is said to be a value that demands respect, and respect for this value is incompatible with making comparisons and trade-offs concerning the good of individuals (Velleman 1999: 370). Because Rational Impartialism holds that one ought to have equal fundamental concern for the equal good of all individuals, this implies that agents have ultimate reasons to prioritize the greater good of some individuals over the lesser good of others.

And this seems to run afoul of the idea that every person has a value that precludes comparisons and trade-offs. Hence, one might worry that an agent who attempts to live up to the standards of RI would alienate herself from her fellow human beings insofar as she would not be directly responsive to the true value of their personhood. This would not only constitute a rational failing on the part of the agent, it could also constitute a significant harm to those near and dear to her.

In order to address this worry we must first note that RI does not require of agents that they respond to everyone's value similarly *in practice*. As mentioned above, the impartiality required by RI concerns ultimate reasons. This is compatible with our having derivative reasons to focus our attention on certain individuals including ourselves and our intimates. In most cases, for instance, parents should show special concern for the welfare of their own children. The same is true of friends, siblings, spouses, etc.

That being said, RI does imply that at the fundamental level, I ought to give equal weight to the equal good of everyone. Displays of partiality are justified if they are impartially optimal. Hence, we must still answer the question of whether all persons possess a type of value that precludes reasons to make comparisons with respect to the good of others.

The suggestion that comparing the good of some against the good of others can be inappropriate in certain contexts is plausible enough. But the thought that persons possess an intrinsic value that precludes making comparisons altogether, if understood literally, is obviously false. Even the strictest Kantian would have to accept the rationality of comparing the good of some against the good of others when deciding what one has most reason to do (Keller 2013: 131). When deciding how to allocate the money I have set aside for charity, it would be absurd not to consider and compare the needs of those individuals who receive help from the respective charities I might donate to. Giving priority to whoever has the greatest need does not betray a

lack of respect for the value of individuals. And even when there is a choice to be made between benefitting someone with whom you have a special relationship and benefitting a complete stranger, it is easy to see that a comparison of the value of the interests is possible and sensible. Keller provides a helpful example in which you must choose between treating your child's asthma and saving several other children from blindness. Even if you believe in ultimate reasons of partiality, and thus believe that you have ultimate reasons to treat your child's asthma rather than saving the other children from blindness, you can at least see that a neutral observer would have reason to choose the latter. And that is all that is necessary to see that comparisons and trade-offs can be legitimate (ibid.).

To the question of whether general acceptance of and adherence to RI would alienate us from one another, it is helpful to imagine a world in which, at the fundamental level, everyone did in fact have equal fundamental concern for the equal good of all. In responding to a version of Williams' integrity objection I noted that because we do not live in an ideal world, most of us living in affluent nations have decisive reasons to change many things about our lives, including the amount of resources we devote to making progress on issues like global poverty. But think for a moment about what things would be like should everyone come to recognize that there are no ultimate reasons of partiality, and that the good of distant strangers provides ultimate reasons of the same strength as the reasons provided by one's own good. Were all citizens of wealthy nations to become convinced of this and attempt to live their lives accordingly, it is doubtful that problems like severe global poverty would persist for much longer. And though we would likely need to maintain certain special relations with certain individuals, if everyone were to share the same basic commitment to treating the good of everyone as an equal source of ultimate reasons we should not expect that this would result in a loss of meaning or connection with those around

us. If anything, universal recognition of the equal normative significance of all should lead to a greater sense of unity and connectedness for everyone.

4 Negative-Sum Altruism

In chapter 1, I framed the Partiality Question as concerning the extent to which self-interest and the interests of one's intimates carry greater rational weight for one than the interests of strangers. If Rational Impartialism is true, the answer to this question is No. Though I claimed that the Partiality Question is the most fundamental and important question of philosophical ethics, RI has significant implications for other important questions. One particularly important question is whether an agent can ever have ultimate reason to perform an act of what I shall call 'negative-sum altruism'. Acts of negative-sum altruism involve sacrificing one's own greater good for the sake of the lesser good of someone else. Imagine, for instance, that an agent is trapped in some wreckage following an earthquake, and that she could save a stranger from losing her leg by sacrificing her own life.¹¹ According to common consciousness, this act would be morally and rationally permissible (assuming that making the sacrifice would not prevent you from meeting important obligations to others). Indeed, many people would likely deem such an act of selflessness praiseworthy. At the same time, however, few would say that this act is morally or rationally required. Hence, the most generally shared judgement would likely be that the act was supererogatory. You had sufficient reason to make the sacrifice, but you also had sufficient reason not to.

Rational Impartialism implies a different verdict about such cases. According to RI, one ought to have equal fundamental concern for the equal good of everyone. Hence, one could not

¹¹ This example is based on examples from Parfit (2011, 1: ch. 8).

have sufficient reason to sacrifice one's lesser good for the greater good of someone else.

Insofar as many people share the intuition that negative-sum altruism can be rationally justified, this is a further respect in which RI deviates from common sense.

In responding to this worry, we can begin by noting that it is unsurprising that we find it intuitive that a person might have sufficient ultimate reason to sacrifice a greater good of hers for the sake of a lesser good of someone else. Given the human tendency to prioritize self-interest, an agent who is willing to prioritize the lesser needs of others ahead of her own greater needs reveals a unique disposition to extend her concern to others to a degree that would typically result in a net increase in impartial good. Such an agent likely possesses the traits of generosity and compassion, which are regarded as among the chief moral virtues.

But although the traits that can manifest themselves in negative-sum altruism are generally beneficial for everyone, the important question is whether an agent could ever have sufficient ultimate reason to perform such an act. In order to get the question clear, we must make sure that we consider only cases that are genuine instances of negative-sum altruism. Some scenarios which might seem to be properly classified as such are actually not because of the further consequences of the act. For instance, if the sacrifice were to become widely known, this might inspire people to show more kindness and generosity to others, which could make it the case that the act lead to a net increase in total good, and hence optimal from an impartial perspective. We must also set aside the point that the traits that manifest themselves in acts of negative sum altruism—generosity, kindness, etc. are typically beneficial from an impartial perspective.

Given these criteria, we can use the earthquake example mentioned above, so long as we stipulate that the agent in position to make the sacrifice knows that nobody will ever know about her decision. So imagine the following scenario:

Earthquake: Two strangers, Mollee and Carmen, are trapped in slowly collapsing wreckage. Though Carmen is unaware of Mollee's presence, Mollee can see that Carmen's leg will be crushed if Mollee does not move some of the debris. But Mollee also knows that the cost of saving Carmen's leg will be certain death from an onslaught of rubble that would be jarred loose. Mollee makes the sacrifice and is subsequently killed. Carmen's leg is saved, and a few hours later she is pulled from the debris by rescue workers. Nobody ever learns of Mollee's sacrifice.

Mollee's decision to sacrifice her life for a stranger's leg is undeniably generous. But the important question is whether it was justified by the light of reason. Rational Impartialism implies that it was not. The reason why it was not justified is that Carmen's good is no more valuable than Mollee's, and as a rational being Mollee ought to give equal weight to equal portions of the good. Mollee's life is a source of more value than Carmen's leg. Hence, there is more counting in favor of Mollee's letting Carmen's leg be destroyed than there is for sacrificing her life. The fact that Mollee's act of negative-sum altruism was generous and driven by a deep concern for the welfare of others is not sufficient to justify making such a choice.

I suspect that some will want to defend acts of negative-sum altruism because such acts are widely held to be morally permissible. And it certainly seems implausible that an agent who sacrifices her own greater good for the lesser good of someone else has done something morally wrong, and hence is deserving of blame. Considerations such as these further support my

preference for avoiding moral concepts such as ‘permissible’ when engaging in theoretical ethics. It is indeed highly plausible that acts such as the one chosen by Mollee in the above example do not merit blame. But that is not to say that they are justified by the light of reason.

After his axiomatic deduction of the maxim of Benevolence in *Methods* III:XIII, Sidgwick claims that, once we set aside the distorting influence of derivative reasons of partiality, his impartialist principle is supported by common-sense. He provides the following remarks:

I think that a ‘plain man,’ in a modern civilized society, if his conscience were fairly brought to consider the hypothetical question, whether it would be morally right for him to seek his own happiness on any occasion if it involved a certain sacrifice of the greater happiness of some other human being,--without any counterbalancing gain to any one [sic] else,--would answer unhesitatingly in the negative. (ME 382)

If we replace ‘morally right’ with the ‘ought’ of practical reason, I believe similar remarks would apply in the case of negative-sum altruism. I suspect that a ‘plain man’ would deny that one ought to seek the happiness of others when this involves a sacrifice of one’s own greater happiness, and there is no counterbalancing gain to anyone else. And if so, this suggests an independent challenge for proponents of partialism. If we hold that it would generally be a mistake to count one’s own good less than the good of others, should we not also hold that it would be a mistake to count one’s own good more? The best explanation for our judgement that one ought not count one’s own good less is that one’s own good is no less valuable than anyone else’s. But nor is one’s own good any more valuable than anyone else’s. So by parity of reasoning, we should conclude that one ought to give equal weight to the equal good of all. At

the very least, the counterintuitiveness of counting one's good less throws an *onus probandi* on those who argue that self-interest carries extra rational weight.¹²

5 Conclusion

My aim in this chapter has been to diminish the air of paradox surrounding Rational Impartialism. I have addressed three general worries that might make it difficult for some to accept the truth of RI: that the view is implausibly demanding; that adhering to it would lead to alienation; and that it runs afoul of common judgements about altruism and supererogation. With respect to demandingness, I argued that pervasive irrationality in the objective-reason sense follows from any plausible theory of practical reason. I argued further that our intuitions about cases involving extreme sacrifice or extreme impartiality do not tell strongly against RI once we bring all the relevant considerations into focus and eliminate possible distorting influences.

In response to alienation worries, I argued that appeals to the importance of an agent pursuing her own personal projects and commitments cannot be sufficient for undermining impartialist views because the mere fact that an agent cares deeply about a cause or project does not imply that its pursuit is justified by the light of reason. I noted further that RI is in principle compatible with agents pursuing the projects and activities that most plausibly contribute to their welfare. I also considered the worry that RI is incompatible with responding appropriately to the value of individuals. In response, I explained why making comparisons and trade-offs concerning the good of persons is both necessary and innocuous, and that giving equal weight to the good of everyone when deciding how to act does not preclude having respect for each individual.

¹² Thanks to Julia Driver for alerting me to this particular challenge for proponents of partiality.

Finally, I argued that intuitions supporting the reasonableness of negative-sum altruism can be explained away by the fact that such acts typically result from traits and motives that are generally beneficial for the greater good. Additionally, we tend to think that such acts must be rationally permissible because they are clearly morally permissible. When we jettison these moral terms and reflect on the simple question of what an agent has most reason to do, it becomes clear that one should not count one's own good less than the equal good of someone else.

This concludes the presentation and defense of my Sidgwickian argument for Rational Impartialism. In describing his masterpiece, Sidgwick writes: "The book solves nothing, but may clear up the ideas of one or two people, a little."¹³ My central aim in writing this dissertation has been to show that Sidgwick provides most of the resources needed for solving what I take to be the most fundamental and important question of philosophical ethics. If we accept that practical reasons are object-given and value-based, then we should also accept that the good of all is a source of reasons for all. And when we compare Sidgwick's axiom of Equal Good with the alternative maxim championed by proponents of partiality, we find that the former is on more stable ground. Hence, we can conclude that there is not, as Sidgwick feared, a fundamental contradiction in one of the chief departments of human thought. Though Sidgwick was right in his observation that each individual has special concern for her own existence, this fact about our inclinations does not track deep ethical truths. When we encounter basic conflicts between personal or partial interests and the greater good of all, our ultimate reasons tell decisively in favor of the latter.

¹³ *Henry Sidgwick: A Memoir* by A.S. and E.M.S. (Macmillan 1906). Reference from Parfit (2011, 1: xxxiv).

References

- Anscombe, G.E.M. (1997). "Modern Moral Philosophy." In *Virtue Ethics*. Roger Crisp and Michael Slote (eds.). Oxford: Oxford University Press.
- Anthony, A. (2014). "Would You Give Your Kidney to a Complete Stranger?" *The Guardian* Sunday 13 July 2014. <http://www.theguardian.com/society/2014/jul/13/kidney-donation-altruism-andrew-anthony>.
- Audi, R. (2010). *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. 3rd edition. New York: Routledge.
- . (2015). "Intuition and Its Place in Ethics." *Journal of the American Philosophical Association* 1: 57-77.
- Bernstein, L. (2015). "It's a Week into January and a Quarter of Us Have Already Abandoned Our New Year's Resolutions." *The Washington Post* Wednesday 7 January 2015. <http://www.washingtonpost.com/news/to-your-health/wp/2015/01/07/its-january-7-and-a-quarter-of-us-have-already-abandoned-our-new-years-resolutions>.
- Blanshard, B. (1984). *Four Reasonable Men: Marcus Aurelius, John Stuart Mill, Ernest Renan, Henry Sidgwick*. Middletown, Conn.: Wesleyan University Press.
- Brink, D. (1997). "Self-Love and Altruism." *Social Philosophy and Policy* 14/1: 122-57.
- Broad, C.D. (1930). *Five Types of Ethical Theory*. London: Kegan Paul, Trench, and Trubner.
- Chappell, R. (2012). "Fittingness: The Sole Normative Primitive." *Philosophical Quarterly* 62/249: 684-704.
- Christensen, D. and J. Lackey. (2013). *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.
- Crisp, R. (2002). "Sidgwick and the Boundaries of Intuitionism." In *Ethical Intuitionism: Evaluations*. P. Stratton-Lake (ed.). New York: Oxford University Press.
- . (2006). *Reasons and the Good*. Oxford: Clarendon Press.
- . (2015). *The Cosmos of Duty: Henry Sidgwick's Methods of Ethics*. Oxford: Oxford University Press.
- D'Arms J. and D. Jacobson. (2000). "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61/1: 65-90.

- de Lazari-Radek, K. and P. Singer. (2014). *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.
- Dorsey, Dale. (2012). "Weak Anti-Rationalism and the Demands of Morality." *Nous* 46: 1-23.
- . (2013). "Two Dualisms of Practical Reason." In *Oxford Studies in Metaethics* Volume 8. Russ Shaffer-Landau (ed.). New York: Oxford University Press.
- . (unpublished). *The Limits of Moral Authority*.
- Driver, J. (2012). *Consequentialism*. New York: Routledge.
- Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Findlay, J. (1961). *Values and Intentions*. London: George Allen and Unwin.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Goodin, R. (1985). *Protecting the Vulnerable*. Chicago: University of Chicago Press.
- Haidt, J. (2001). "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review*, 108: 814-34.
- Hurka, T. (2014). "Sidgwick on Consequentialism and Deontology: A Critique." *Utilitas* 26: 129-52.
- Irwin, T. (2009). *The Development of Ethics*, vol. 3: *From Kant to Rawls*. Oxford: Oxford University Press.
- Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
- Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.
- Kahane, G. (2011). "Evolutionary Debunking Arguments." *Nous* 45/1: 103-25.
- . (2014). "Evolution and Impartiality." *Ethics* 124: 327-41.
- Kavka, G. (1986). *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.
- Keller, S. (2013). *Partiality*. Princeton: Princeton University Press.
- Kelly, T. (2005). "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology* Volume 1. John Hawthorne and Tamar Gendler (eds.). New York: Oxford University Press.

- Kraut, R. (2011). *Against Absolute Goodness*. New York: Oxford University Press.
- McHugh, C. and J. Way. (forthcoming). "Fittingness First." *Ethics*.
- McElwee, B. (2007). "Consequentialism, Demandingness, and the Monism of Practical Reason." *Proceedings of the Aristotelian Society* 107: 359-74.
- . (2010). "Consequentialism and Permissibility." *Utilitas* 22 (2): 171-83.
- McLeod, O. (2001). "Just Plain 'Ought'." *Journal of Ethics* 5 (4):269-91.
- Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.
- Marquis, D. (1989). "Why Abortion is Immoral." *Journal of Philosophy* 86 (4): 183-202.
- Mill, J.S. (1998). *Utilitarianism*. Roger Crisp (ed.). Oxford: Oxford University Press.
- Moore, G.E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, T. (1986). *The View from Nowhere*. Oxford: Oxford University Press.
- Norcross, A. (2008). "The Scalar Approach to Utilitarianism." In *The Blackwell Guide to Mill's Utilitarianism*. Henry West (ed.). Malden, MA: Blackwell.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Parfit, D. (2011). *On What Matters*. Oxford: Oxford University Press.
- Parfit, D. (unpublished). "How We Can Avoid the Repugnant Conclusion."
- Phillips, D. (2011). *Sidgwickian Ethics*. Oxford: Oxford University Press.
- Plato. (1997). *The Republic*. In *Plato: Complete Works*. J.M. Cooper (ed.). Indianapolis: Hackett.
- Portmore, D. (2011). *Commonsense Consequentialism*. New York: Oxford University Press.
- Prinz, J. (2009). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Rabinowicz, W. and T. Ronnow-Rasmussen. (2004). "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114: 391-423.
- Railton, P. (1984). "Alienation, Consequentialism, and the Demands of Morality." *Philosophy & Public Affairs* 13: 134-71.
- Rashdall, H. (1907). *The Theory of Good and Evil: A Treatise on Moral Philosophy*. Oxford: Clarendon Press.

- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Raz, J. (2011). *From Normativity to Responsibility*. Oxford: Oxford University Press.
- Scanlon, T.M. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Scheffler, S. (1982). *The Rejection of Consequentialism*. Oxford: Clarendon Press.
- Schneewind, J. (1977). *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Clarendon Press.
- Schopenhauer, A. (1995). *On the Basis of Morality*. Indianapolis: Hackett.
- Shafer-Landau, R. (2010). *The Fundamentals of Ethics*. New York: Oxford University Press.
- Shaver, R. (1999). *Rational Egoism: A Selective and Critical History*. New York: Cambridge University Press.
- . (2014). "Sidgwick's Axioms and Consequentialism." *Philosophical Review* 123: 173-204.
- Sidgwick, A. and E.M. Sidgwick. (1906). *Henry Sidgwick: A Memoir*. London: Macmillan.
- Sidgwick, H. (1874). *The Methods of Ethics* 1st edition. London: Macmillan.
- . (1981). *The Methods of Ethics* 7th edition. Indianapolis: Hackett.
- . (2000). *Essays on Ethics and Method*. Marcus Singer (ed.). Oxford: Clarendon Press.
- Singer, P. (1981). *The Expanding Circle*. Princeton, NJ: Princeton University Press.
- Skarsuane, K. O. (2011). "Darwin and Moral Realism: Survival of the Iffiest." *Philosophical Studies* 152: 229-43.
- Skelton, A. (2008). "Sidgwick's Philosophical Intuitions." *Etica e Politica/Ethics and Politics* 10: 185-209.
- Skorupski, J. (1999). *Ethical Explorations*. Oxford: Oxford University Press.
- Slote, M. (1983). *Goods and Virtues*. Oxford: Oxford University Press.
- Sobel, D. (unpublished). "Why Not Subjectivism?"
- Sober, E. and D.S. Wilson. (1998). *Unto Others*. Cambridge, Mass: Harvard University Press.
- Strawson, P.F. (1962). "Freedom and Resentment." *Proceedings of the British Academy*. 48: 1-25.

- Street, S. (2006). "A Darwinian dilemma for realist theories of value." *Philosophical Studies* 127 (1): 109-166.
- . (2012). "Coming to terms with contingency : Humean Constructivism about Practical Reason." In *Constructivism in Practical Philosophy*. Jimmy Lenman and Yonatan Shemmer (eds.). Oxford: Oxford University Press.
- . (2015). "Does Anything Really Matter or Did We Just Evolve to Think So." In *The Norton Introduction to Philosophy*. Gideon Rose, Alex Byrne, Joshua Cohen, and Seana Shiffrin (eds.). New York: W.W. Norton & Company.
- Stroud, S. (1998). "Moral Overridingness and Moral Theory." *Pacific Philosophical Quarterly* 79 (2): 170–189.
- Temkin, L.S. (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.
- Velleman, J. D. (1991). "Well-Being and Time." *Pacific Philosophical Quarterly* 72 (1): 48-77.
- . (1999). "Love as a Moral Emotion." *Ethics* 109 (2): 338-374.
- Williams, B. (1973). "A Critique of Utilitarianism." In *Utilitarianism: For and Against*. J.J.C. Smart and Bernard Williams. Cambridge: Cambridge University Press.
- . (1981). "Persons, Character, and Morality." In *Moral Luck*. James Rachels (ed.). Cambridge: Cambridge University Press