Topics in Rational Choice Theory: 
Altruism, Consequentialism, and Identity

by

Sahar Zahida Akhtar

Department of Philosophy
Duke University

Date: ________________________

Approved:

___________________________
Allen Buchanan, co-supervisor

___________________________
Alex Rosenberg, co-supervisor

___________________________
Jesse Prinz

___________________________
Susan Wolf

___________________________
David Wong

Dissertation submitted in partial fulfillment of 
the requirements for the degree of Doctorate 
of Philosophy in the Department of 
Philosophy in the Graduate School 
of Duke University

2008
ABSTRACT

Topics in Rational Choice Theory:
Altruism, Consequentialism, and Identity

by
Sahar Zahida Akhtar
Department of Philosophy
Duke University

Date:

Approved:

___________________________
Allen Buchanan, co-supervisor

___________________________
Alex Rosenberg, co-supervisor

___________________________
Jesse Prinz

___________________________
Susan Wolf

___________________________
David Wong

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctorate in Philosophy in the Department of Philosophy in the Graduate School of Duke University

2008
Abstract

Rational Choice theory includes a broad body of research that attempts to account for how people act in a variety of contexts, including economic, political and even moral situations. By proposing, most generally, that individuals rationally pursue their self-interests regardless of the context, rational choice has had extensive theoretical and empirical success, on the one hand, and has also faced wide criticism when applied in a variety of disciplines, on the other hand. While there is disagreement over what the defining assumptions of rational choice theory are, in this dissertation I focus on three on which there is widespread agreement. These three features of rational choice theory are: its assumption of egoism or self-interest as the central motivation of individuals; its reliance on consequences as part of a comparative decision-making framework; and finally, its focus on the individual and not on groups as the methodological and normative unit of analysis.

In correspondence to these three features, my dissertation is divided into three parts and explores the separate topics of (I) egoism and altruism; (II) consequentialism and ethical decision-making; and, (III) individualism and group identity. The dissertation is not an exercise in showing the extensive problems of rational choice theory, although there are many. The dissertation rather engages these three topics with differing results, some of which in fact attempts to revitalize rational choice, or at least features of rational choice. For the part on altruism, my goal is to demonstrate why the central assumption of egoism in rational choice theory is problematic. More broadly, I argue for a different way of defining genuine altruistic motivation. A result of my analysis there is that altruism
appears to be more widespread than has been traditionally assumed and is more amenable to empirical examination. For my discussion on consequentialism, my aim is to re-characterize rational choice as a mode of moral decision-making. I argue that the moral agent is one who frequently compares her particular moral ends in a stable fashion and for this reason cost-benefit analysis is a fully moral framework, one that encourages the agent to genuinely care for her ends and values. For the topic of individualism and group identity, my objective is to show how a previously dismissed topic, once unpacked, is fully consistent with rational choice theory and ought to be of interest to the rational choice theorist. I show that if the liberal political theorist, including the rational choice theorist, is to value group identity, the commitment is only limited to valuing a form of group identity—particularized identity—that is individualist in character.
Contents

Abstract.......................................................................................................................iv
List of Tables...............................................................................................................ix
List of Figures..............................................................................................................x
Introduction ................................................................................................................1
  Altruism...............................................................................................................3
  Consequentialism.................................................................................................8
  Identity..............................................................................................................14
Part I: Rational Choice and Altruism
  1. Genuine Altruism: What is it and Does it Exist..............................................22
     1.1 Introduction.................................................................................................22
     1.2 Standard View of Altruism.........................................................................26
     1.3 Evolutionary Arguments...........................................................................31
     1.4 Problems with Evolutionary Arguments for Genuine Altruism.............40
     1.5 Conclusion................................................................................................48
     1.6 Appendix 1: Are Feelings and Desires Separate?.................................51
  2. Ultimate or Instrumental Altruism? Who Cares............................................64
     2.1 Introduction.................................................................................................64
     2.2 Prior Suitability..........................................................................................65
     2.3 Central Role of Feelings...........................................................................67
        2.3.1 Reason I...............................................................................................68
        2.3.2 Reason II.............................................................................................73
        2.3.3 Reason III...........................................................................................87
     2.4 Conclusion................................................................................................97
2.5 Appendix 2: Applications to Rational Choice ........................................ 99
   2.5.1 New Avenues for Theoretical Research ........................................ 99
   2.5.2 New Avenues for Empirical Research ........................................ 108

Part II: Rational Choice and Consequentialism

3. The Structure of Moral Decision-Making ............................................. 115
   3.1 Introduction .................................................................................... 115
   3.2 Against General Value .................................................................. 118
   3.3 Mediating Ends in Our Everyday Lives ......................................... 127
   3.4 Considering Each of Our Ends ...................................................... 136
   3.5 Mediating Our Moral Values ....................................................... 142
   3.6 Conclusion .................................................................................... 148

4. Cost-Benefit Analysis in Ethical Decision-Making .............................. 151
   4.1 Introduction .................................................................................... 151
   4.2 Intrinsic Value and Alienation ...................................................... 152
   4.3 Comparing Values in Bioethics .................................................... 161
   4.4 Beyond Alienation: Other Costs and Benefits of Using CBA ............ 169
   4.5 Conclusion .................................................................................... 175

Part III: Rational Choice and Individualism

5. Liberal Respect for Identity? Only for Particular Ones ...................... 177
   5.1 Introduction .................................................................................... 177
   5.2 Identity and Autonomy .................................................................. 180
   5.3 Including Identity in Respect for Persons ..................................... 186
   5.4 The Locality of Respect ............................................................... 189
   5.5 Collective vs. Particularized Identity ............................................ 197
5.6 Practical Implications of the Debate........................................202

5.7 Conclusion..............................................................................209

Bibliography.................................................................................211

Biography......................................................................................217
List of Tables

Table 1 ........................................................................................................51
List of Figures

Figure 1 ..............................................................................................................38
Figure 2 ..............................................................................................................44
Figure 3 ..............................................................................................................74
Figure 4 ..............................................................................................................80
Figure 5 ..............................................................................................................92
Introduction

Rational Choice theory includes a broad body of research that attempts to account for how people act in a variety of contexts, including economic, political and even moral situations. By proposing, most generally, that individuals rationally pursue their self-interests regardless of the context, rational choice has had extensive theoretical and empirical success, on the one hand, and has also faced wide criticism when applied in a variety of disciplines, on the other hand. While there is disagreement over what the defining assumptions of rational choice theory are, in my dissertation I focus on three on which there is widespread agreement. These three features of rational choice theory are: its assumption of egoism or self-interest as the central motivation of individuals; its reliance on consequences as part of a comparative decision making framework; and finally, its focus on the individual and not on groups as the methodological and normative unit of analysis.

In correspondence to these three features, my dissertation is divided into three parts and explores the separate topics of (I) egoism and altruism; (II) consequentialism; and, (III) individualism and group identity. The dissertation is not an exercise in showing the extensive problems of rational choice theory, although there are many. The dissertation rather engages these three topics with differing results, some of which in fact attempts to revitalize rational choice, or at least features

---

of rational choice. For the part on altruism, my goal is to demonstrate why the central assumption of egoism in rational choice theory is problematic. However, for my discussion on consequentialism, my aim is to re-characterize rational choice as a mode of moral decision making. For the topic of individualism and group identity, my objective is to show how a previously dismissed topic, once unpacked, is fully consistent with rational choice theory and ought to be of interest to the rational choice theorist.

The emphasis on these topics is distributed according to how extensively each topic engages the rational choice literature. Thus, I spend the most significant portion of my dissertation, Chapters 1 and 2, with their subsequent Appendices, discussing altruism because this issue has been part of a longstanding debate surrounding rational choice theory—the view that there is no genuine altruism is an essential feature of the rational choice tradition. I spend relatively less time on the topic of consequentialism. While consequentialism is an important feature of rational choice, in Chapters 3 and 4, I demonstrate how a different defense of consequentialism, based on what it means to really care about ones moral ends, can offer us an alternative view of the nature of rational choice—not as a systematic descriptive and normative theory but as a mode of decision making in ethics. Finally, the topic of group identity is often not coupled with the rational choice literature. However, in Chapter 5, I show how a unique, and individual-based, understanding of identity allows for its inclusion in the broad rational choice paradigm.

In addition to different topics, my dissertation also spans three different fields of philosophy: moral psychology in part I, normative and applied ethics in part II, and
political philosophy in the final part. Thus it has the virtue of connecting these
disparate areas of philosophical inquiry into an overall project on rational choice.
Each of the subsequent two topics relates to the first part on altruism in various ways.
For instance, the discussion of consequentialism involves what it means to really care
for one’s particular ends, and has affinities with the discussion of what it means to
really care for others in the context of the debate between egoism and altruism, and
the discussion on identity relates to altruism because identity is often conflated with
altruism. In what follows, I describe each of these topics in greater detail as well as
illuminate some connections between them and explain how these connections as
well as the dissertation more generally will serve as the basis for future work.

Altruism

In Chapters 1 and 2, I discuss the common assumption of egoistic, or self-
interested, motivation in rational choice theory. My analysis does not directly purport
to show that this assumption is incorrect and that altruistic motivation is widespread,
but it does so indirectly. I try to demonstrate that the standard framework for
conceiving of altruism and egoism is misguided. Specifically, I argue that the
standard conception of ‘genuine’ altruistic motivation is insufficient because it omits
any reference to feelings. I then argue for a different way of thinking about genuine
altruistic motivation, one in which feelings have a privileged role. A result of my
analysis is that there seems to be a continuum of motivation, ranging from egoistic to
altruistic, and that the altruistic sort appears to be more widespread than has traditionally been recognized.

Both the rational choice and philosophy literature describe altruistic motivation in terms of a desire or preference for the improved welfare of another. Combined with a relevant belief, such as the belief that someone has experienced harm, a desire or preference for the improvement or protection of another’s welfare constitutes altruistic motivation. *Genuine* altruistic motivation (genuine altruism, henceforth) is furthermore defined in terms of an ultimate desire or preference for the improved welfare of another, one that cannot be reduced to any desires for the improvement of one’s own welfare.\(^2\) On the other hand, if someone possesses a desire to help another but the satisfaction of that desire only serves as a means to the satisfaction of a desire for one’s own welfare this would be a case of reductive, or instrumental, and not genuine, altruism. If the desire to improve one’s welfare is ultimate, then this would constitute egoism.\(^3\)

To see the contrast between egoism and genuine altruism more clearly, consider an example. If I have a desire to give money to charity, this is an altruistic motivation. But we have not yet determined whether it is genuine altruism or reductive altruism. According to the standard view, it would be genuine altruism if

---


my desire to give money is an ultimate one—if I have no other desire to give money to charity other than that it would help others. On the other hand, if I desire to give money to charity only because I ultimately desire the feeling that I have made a difference in the world and giving money is a way to make a difference in the world, then my desire to give money would be a case of reductive altruism, which according to the standard view, would still constitute egoism. Egoism can consist either in the pursuit of any kind of external good, such as material goods only or material goods combined with public approval, distinction or fame, or can involve the pursuit of pleasure and other internal rewards—the latter being hedonism, one specific form of egoism.4

While the question of whether there is any altruistic motivation is uncontroversial—it seems obvious that people do have desires to help others or, more generally, desires for the improved welfare of others—the answer to the question of whether there is genuine altruistic motivation is far less obvious. That is, at least if we maintain the standard definition of genuine altruistic motivation. This is because a persistent problem for the existence of genuine altruism has been that critics can always seemingly reduce an altruistic motivation to seemingly egoistic motivations—to prudential motivational explanations such as “she’s kind to them because otherwise she would not have relationships with them,” or, especially, to hedonistic motivational explanations such as “he gives to charity because it feels good to give to charity,” and “he feeds his children because it hurts to see one’s children go hungry”.

Arguments for the existence of egoism consisting of external rewards can be easily refuted by citing examples of people taking action where there are no material or recognition returns to be had (a prime example is voting in elections). Therefore hedonism provides the strongest position for the thesis of egoism.

Since internal rewards are opaque to researchers and since egoism is a robust descriptive thesis about many of our motivations, proponents of reductive altruism argue that there simply is no need to posit genuine altruism. Researchers who take a neutral position on the question have noted, as Daniel Dennett does, that “genuine, or pure, altruism is an elusive concept, an ideal that always seems to evaporate just when you get in position to reach out to grab it. It isn’t clear what would count as genuine altruism, and paradox hovers constantly nearby.”

A variety of empirical, philosophical and evolutionary arguments have been advanced to demonstrate the existence of genuine altruism. However, in the first two chapters of my dissertation, I attempt to show why all of these arguments are unsuccessful. In Chapter 1, I demonstrate that the most plausible evolutionary argument, put forth by Sober and Wilson, fails. Because they provide a novel argument for the widespread existence of genuine altruism and because an evolutionary framework shares important features with rational choice theory, I spend some time on Sober and Wilson’s argument. I then demonstrate the failure of both the philosophical arguments and the empirical work that attempt to show that genuine altruism exists.

---

In each of these lines of research, I argue in the second chapter, the source of the failure to establish the existence of genuine altruism is the way in which genuine altruism is defined. Specifically, I argue that while we often think of altruistic motivation in terms of feelings such as sympathy and empathy, the standard definition of genuine altruism makes no essential reference to these feelings. Rather, the standard account makes the presence of an ultimate altruistic desire or preference the essential feature of genuine altruism. I argue that, for a variety of reasons, whether or not a desire or preference is ultimate does not satisfactorily address whether the motivation is genuine. I then argue that the presence of certain kinds of feelings better indicates genuineness in motivation. Furthermore, I argue that feelings and desires/preferences are distinct and that feelings can be motivating. My goal is not to specify the range of feelings that constitute genuine altruism, but rather to show that certain types of feelings are essential to our conception of genuine altruism.

My analysis of the central importance of feelings to genuine altruistic motivation has broad implications for rational choice theory and its application to explain behavior. First, my proposed account of genuine altruism clears away the conceptual baggage that has notoriously prevented both an acceptance that genuine altruism exists and efforts to explore the reach and contexts of altruistic motivation. Second, once we account for the prominent role of feelings in understanding genuine altruism, and at the same time recognize the limited, and perhaps absent, role of ultimate desires, several avenues for empirical research are opened up. Finally, whereas the standard view of altruism does not always seem to illuminate some forms of behavior, a view that privileges feelings may offer insight. I argue that feelings are
separate from desires and preferences and cannot be reduced to them. At the same time, I maintain that feelings can be motivating but that they do not aim to be realized or satisfied in the way that desires do. Because they do not have the aim to be satisfied, they allow us to make better sense of behavior that is expressive rather than consequential or instrumental—expressive behavior just is behavior that does not always seek to achieve any outcome, at least at any given point or in any given circumstance. Altruistic feelings, by causing expressive behavior, therefore seem to have many applications to domains of social, ethical and political life involving community, identity, and social movements.

**Consequentialism**

In Chapter 3 and 4, I focus on another feature of rational choice theory: its emphasis on consequences as the locus for determining whether a particular action, policy or institution is good or bad. There are two ways in which consequences feature in rational choice theory. First, consequences, and in particular wellbeing measured by some form of utility or preference-satisfaction, explicitly figure in welfare economics. Welfare economics is the branch of rational choice theory dedicated to understanding what economic policies or actions make people better off. Importantly, whether or not someone is better off is determined by the relevant consequences. Second, and more subtly, consequences figure in the approach of rational choice. The rational choice theorist believes that to ask whether an action or policy is a good one is unhelpful unless we know what the alternative to that action or policy is. Rational choice therefore is the process of choosing the best, however best
is to be defined, selection among available options. Furthermore, the rational choice theorist believes that to ask whether one action or policy is better than another is already to imply a common measurement from which comparisons can be made. By being comparative in analysis, rational choice then places the emphasis on consequences, and specifically, on general values. This is because, as I will demonstrate in the third chapter, general values can more easily serve as the basis from which to make comparisons between competing options than can particular ends.

When it comes to value theory—to our moral options—many moral theorists ranging from Kantians to virtue ethicists to particularists, deny that consequences ought to serve as the only or main determinant of our values, or of whether an action ought to be taken or a disposition nurtured. In Chapters 3 and 4, rather than giving a defense of consequentialism based on its purported rationality or coherence, I try to provide a defense of consequentialism based on what it means to really care for one’s ends.

There is a widespread belief among philosophers that a genuine care or concern is one that does not reduce to other concerns. A genuine care or concern is one that has intrinsic, and not merely instrumental, value. Indeed, as discussed above, it is this belief that has shaped the definition of genuine altruism, of sharply distinguishing altruism from any form of egoism. Philosophers have argued that serious psychological and ethical tensions arise when we are ultimately concerned with values of a general or abstract nature, rather than with our particular ends, relationships, and projects. For instance, Kantianism and Utilitarianism, with their
groundings in duty and maximization of happiness, respectively, are criticized for being too impersonal, of alienating us from our own personal relationships and plans, and of leading to a loss of integrity, that is, of the value of being specially responsible for and acting on one’s deeply held dispositions and motives. Most recently, Michael Smith argues that the morally good person is one who is motivated directly, and non-instrumentally, by her particular moral ends. Just as good lovers must have a direct concern for their loved ones, good people, Smith argues, must have a direct concern for equality, justice, honesty, the welfare of people, and other things that they believe are right.

In Chapter 3, I argue that many of us implicitly try to make decisions about our ends and relationships in the very way which is charged with being problematic. Furthermore, I argue that trying to mediate conflicts between one’s particular moral ends, such as when one is faced with helping the distant poor or protecting the rights of immigrants, by appealing to more general values, such as the promotion of wellbeing, is on reflection how we tend to think about the morally good person. Appealing to general values that are separate from particular ones can encourage broad reflection about all of one’s particular ends, including how they fit together, and can reveal their shared importance.

---


While relying on general values to sort through conflicts does not necessarily entail that these more general values have specific consequentialist content, as described above it does reflect the more subtle reliance on consequentialism as a mode or approach of ethical thinking; the mode is a comparative one. Furthermore, I argue with the rational choice theorist that even the question of whether general values can help us settle conflicts between our particular values and ends in a morally satisfying way is itself a comparative exercise. The question is not whether an appeal to general values fully solves the moral problem of conflict in one’s ends, but whether there are good reasons to think that appealing to general values does so better than alternative approaches.

In Chapter 4, I extend my discussion of consequentialism to defend the use of cost-benefit analysis in ethics. Cost-benefit analysis does not always assume that the ends or values to be compared are consequences, but does possess the mode of ethical thinking I described above. I argue that while traditional cost-benefit analysis that relies on some notion of efficiency, such as pareto efficiency, is subject to a wide range of moral objections, cost-benefit analysis as conceived of a framework in which to compare our moral ends encourages broad reflection and the consideration of each alternative value. Furthermore, I argue that far from calling for the maximization of any value, cost-benefit analysis simply calls for the minimization of costs, or more minimally, the assessment of costs—and costs are not strictly monetary ones but are rather the values and ends which are not favored whenever an action is undertaken for the sake of some other end or value. The context for my discussion on the use of cost-
benefit analysis in ethics is bioethics, where claims about intrinsic values, and hence the seeming incomparability of values, abound.

Critics of both consequentialist theories and cost-benefit analysis charge that in such theories the value of one’s particular ends reduces to the value of other, more general ends and this implies that consequentialists do not have the appropriate attitude or commitment to their particular ends. The appropriate commitment entails treating one’s ends as intrinsic or ultimate values. My response in these chapters thus shares a general kind of argument with my response in Chapters 1 and 2 that has relevance for rational choice and ethics, and it is the following: whether a motivation, in the context of the debate between egoism and altruism, or a value, in the context of ethical decision making, is an ultimate one does not always in itself tell us very much about the genuineness of that motivation or commitment to that value, respectively. Whereas in the first two chapters I argue that the description of altruism includes references to feelings, in these chapters the focus is not on feelings but rather on the comparative mode of ethical thinking. In each case, however, this means that whether a care or concern is ultimate is not necessarily the critical question to pose.

In future work, I will examine a more explicit relationship between my analysis in the first and second parts of my dissertation by exploring what is required of us by treating altruism as value, and not only as a description of our motivations. Although altruism has received extensive attention in many fields, including economics, anthropology, psychology, and evolutionary biology, it has received comparatively less in ethical discourse, and this is especially true when we consider altruism as a normative value. The differences between normative and descriptive
altruism are many, but to begin with I offer a crude, yet important, one: while it is perfectly consistent with descriptive altruism to love and care for one’s own children and dismiss everyone else in the world, this is not consistent with normative altruism, or its most closely related virtue—compassion.

Compassion has been given little attention in Western secular ethics, relative to Eastern philosophy. Both Adam Smith and David Hume of course recognized our tendencies for fellow-feeling and for perceiving the suffering of others through our ability to sympathize, but it is often argued that their notion of sympathy is much more akin to the way we use the term empathy today, as connoting a capacity for experiencing what others experience and identifying with others—and whether empathy then causes a caring or concern for others is unclear. However we parse out the concepts of empathy vs. sympathy, it can plausibly be argued that neither Smith nor Hume made caring for others central to their normative theories. Other, more contemporary Western ethical theories, including Utilitarianism, demand that one do what is right or virtuous, but there is little emphasis on the imperative to be caring or compassionate, or, even the most closely related virtue, benevolent. Benevolence can be defined, according to Philippa Foot, as having “the proper end… the good of others”.

It might seem that there is a natural fit for benevolence in contemporary virtue or care ethics. Unsurprisingly, something like benevolence is central to care ethics in particular. But the most notable care ethics view, that of Carol Gilligan, is too limited in terms of its ethical requirements on us. Gilligan’s view emphasizes

---

the uniqueness of relationships, with little to no ethical space for caring for others with whom we do not enter relationships. Outside of care ethics, while benevolence is considered to be an important virtue or duty under many views, it is generally one among many.

Rather than advocating a particular ethical view that locates compassion at its center, in future work I will argue that compassion is central to morality in general and that my proposed way of describing altruism and consequentialism is consistent with the requirements of making compassionate decisions, and more generally, of being compassionate. In particular, I will argue that a morality based on compassion is not one that emphasizes the intrinsic or ultimate value either in a variety of separate concerns or the virtues, as virtue ethics can be said to do, or in actions or the good will, as Kantianism does. Rather, compassion-based morality must be broadly consequentialist, and even more, welfarist, and must involve frequent balancing, comparing and weighing of values because of the nature of the world in which we live.

Identity

The final part of my dissertation, Chapter 5, turns to political philosophy to address an apparent tension between rational choice theory, with its close theoretical ties with the liberal tradition, and the assignment of special status to social or group identities, such as religion, culture, and race. The liberal conception of respect for persons, based on the Kantian idea, has been criticized by communitarian authors
such as Michael Sandel and Charles Taylor because, they argue, it fails to include a rich and accurate account of persons. These critics argue that it is not only the universal features of persons that are deserving of respect. Our distinct social identities are also very important—identities that are largely beyond one’s self-reflection and that are highly interdependent on others. Because of the importance of identity, communitarians argue that liberal democratic societies have an obligation to respect the identities of different minority and cultural groups and the failure to do so amounts to a denial of equal respect for persons who are members of those groups. Bhikhu Parekh puts this claim succinctly when he writes ‘the liberal is in theory committed to equal respect for persons. Since human beings are culturally embedded, respect for them entails respect for their cultures and ways of life.’

The overwhelming focus in the debate on identity has been on membership, whether formally or merely associative, in identity groups. The question of whether groups that are formed around shared characteristics ought to be able to collectively pursue their interests or collectively exercise their rights, has been not been an especially contentious matter in liberal discourse. Much more contentious is whether individuals should be accorded special rights or treatment as members of identity-groups, rather than merely as persons or citizens, or whether an individual’s

---

10 This is also roughly the view of other prominent political theorists. For instance, writing in regard to culture, Kymlika argues that the liberal principle of equal respect for persons requires the recognition of the rights of certain cultures since culture is considered “a constitutive part of who the person is” (Kymlika, p. 175) and Taylor argues likewise that an important condition of respecting someone is recognizing that person as being constituted by her culture. (Taylor, Charles. 1992. Multiculturalism and the Politics of Recognition. Princeton: Princeton University Press)

11 As Appiah notes, these collective rights “tend to have more friends…Most people think that it is just fine that Utah or the city of Cambridge or the Catholic church can exercise rights, through the ballot box or (in the case of churches) through whatever consensual internal mechanisms they agree upon.” Appiah, Anthony (2005) The Ethics of Identity. Princeton: Princeton University Press, p. 72-3
status as a member of an identity-group should ever trump an individual’s rights, status, or obligations as a person.\textsuperscript{12} Because identity is almost always viewed in terms of group-membership, valuing identity seems to be at odds with the principles of liberalism and rational choice theory where persons are the ultimate source of value. In particular, the apparent tension between identity recognition and rational choice emerges from the individualism that is prominent in rational choice theory.

Beginning with Max Weber, proceeding with Friedrich Hayek, and later reinforced in the work of game theory, rational choice theorists have asserted the primacy of individual-level over collective-level explanations of phenomena.\textsuperscript{13} Methodological individualism is the view that all actions, even those that result in the collective decisions of societies, associations, businesses, and families, must be explained by reference to the intentional states of individual people. It is important not to confuse this doctrine with the kind of social atomism often associated with Hobbes and which can be described as the view that all of human psychology can be explained pre-socially or without reference to interaction or engagement with other individuals. Methodological individualism does not make any assumptions about the

\textsuperscript{12} It is true that for Kymlicka in particular, the reason we must respect identity-groups such as cultures is because membership in identity-groups provides a context of choice for the individual—meaningful options for a life-plan and a background against which to shape one’s values, norms, and desires. Thus, cultures for instance are important ultimately because individual choice and autonomy matter. However, under Kymlicka’s view individuals would still be granted special rights as members of groups. Members of transitional or waning cultures, he argues, are disadvantaged as compared to the majority cultures with respect to a stable context of choice, and so must be accorded special rights. (Kymlicka, \textit{Liberalism, Community and Culture}.)

content of the intentional states of individuals and is fully compatible with the claim that individual psychology is formed through engagement with others. Since rational choice theory only depends on methodological individualism and not on atomism, it therefore remains open to emphasizing the importance of social identity on how individuals construe themselves and how they conceive the world—that is, as long as we do not describe that social identity as being irreducibly collective in nature.

Methodological individualism has become closely aligned with the stronger thesis of normative individualism, or what Pettit calls ‘personalism’, that is prominent in the tradition of liberalism. Normative individualism, as the name implies, is a doctrine about what constitutes the good. It maintains that whatever is of value about nations, cultures, races, families or any other group, is something that is valuable for the individuals they affect. In short, the thesis denies that there are irreducibly collective goods. Although rational choice theory is not always explicitly committed to normative individualism, there is a wide and significant area of rational choice—welfare economics—that is at least implicitly committed to the liberal thesis of normative individualism. Welfare economics is the branch of economics that is concerned with evaluating different institutions, actions, and policies, on the basis of normative judgments. The single most important analytical tool in this area has been the condition of pareto efficiency, which maintains that a state of the world, X, is

---


15 Goodin argues that the claim that some goods (for instance, cultures, norms and language) are only possible in a society—a non controversial claim—does not show us that once these goods exist their value is not the value for individual persons but rather emergences from some new property of the group or collective itself. Goodin, R. E. 1990. ‘Irreducibly social goods—Comment I’ in *Rationality, Individualism and Public Policy*, ed. G. Brennan and C. Walsh. Canberra: Australian National University.
morally better than another state of the world, Y, if in X at least one person is better-off and no one is worse-off than in Y. The pareto condition is clearly a fundamentally individualist account of the good because “the moral desirability of alternative states of the world is exhausted by the well-offness of persons: any moral goodness is moral goodness for someone.” 16 Thus, a large part of rational choice theory would deny that there is value in social identity if identity is described as being a property of a group or collective itself.

My arguments in the final chapter, however, demonstrate that there is an important way in which we can value social identity that is compatible with both forms of individualism. I begin Chapter 5 by arguing that any liberal account of valuing social identity must be constrained by the liberal understanding of autonomy, which is the broadly Kantian and Rawlsian view that individuals have a capacity to value and to determine their conceptions of the good. I argue that if liberalism is to value identity, it would be subject to ensuring that autonomy is not compromised. Even though placing importance on autonomy is not necessarily a principle of rational choice theory, I will show how satisfying the condition of autonomy will only support valuing a form of social identity that is individualist. In particular, I show that we can accord value and respect to individuals’ identities without ascribing any special status to groups. Thus, even though the explicit starting point of the chapter is liberalism, the result is to countenance a form of social identity that is also fully consistent with both the methodological and normative individualism present in rational choice theory.

The topic of social identity is related to the earlier chapters on altruism because, as mentioned before, identity is often conflated with altruism. As I discuss in my chapters on altruism, altruism refers to the motivation to care for others. Since empathy is a feeling often associated with caring for others and empathy is also closely related to the notion of identifying with others, it might seem that identity is a kind of altruism. However, clear distinctions between altruism and identity emerge in my dissertation. In particular, that I identify with someone else does not yet tell us whether I care about, or am motivated by, that person’s wellbeing. In addition, it seems possible to care about someone without identifying with them in any way, such as when one cares about other animals or even the environment. Furthermore, as we see in Chapter 5, when commentators discuss identity they are often referring to the idea that culture, ethnicity, or race in some important sense constitutes our sense of who we are. These things, they argue, are a significant part of one’s self-conception. However, it is not even clear whether it is specific others that one identifies with or whether it is the norms, values, traditions, and languages particular to a way of life. It seems that at least some of the time it is the latter that is relevant in identity discussions and in these cases, it seems identity has nothing to do with altruism. Conversely, it seems plausible that some times when people express concerns over identity it is really a case of group altruism. For instance, if someone says that all Pakistanis are stupid, I may feel hurt. But it is not clear whether I am experiencing hurt because I am a Pakistani and the negative comments are an attack on who I am, or whether my feeling hurt is for others I care about who happen to be Pakistani. Of course it is possible that it is a little of both, but at least some of the claims made on
behalf of identity may in fact only reflect that we care about others who happen to be like us in some relevant sense. In general, I think part of the reason for the conflation of identity with altruism is because commentators fail to make another distinction—that between individual and group identity. Once we do so, as I do in Chapter 5, we find that at least in certain contexts it is easier to distinguish between identity and altruism. In future work, I will go on to explore the relationship between identity and altruism in the context of caring about distant others with whom one does not identify in any meaningful sense.

In addition to this future work, I will also continue to engage in the conceptual work of distinguishing between not only identity and altruism, but also between identity and ultimate values, as per my discussion in the second part of my dissertation. Writers such as Sandel and Williams speak of the deep commitments that one has as being those things which constitute one’s identity. They suggest that this relationship between identity and deep commitments accounts for why it is alienating to treat one’s commitments as instrumental—that is, non-ultimate—values. But there are some important distinctions between what one values and what constitutes one’s identity. Art and a deep passion for traveling can be things that we value very much without these things being part of anyone’s identity. At the same time, we can have features of our identity that are trivial or even shameful—while we can consistently and perpetually be ashamed of something that is a part of us, it seems the same does not hold for something we value; we would come to no longer value it. Thus, while it may be important to distinguish between what Sandel calls mere preferences and deep commitments, it also seems important to keep distinct the idea
that something is a deep commitment from the idea that something is a constituent of who a person is. While it seems plausible that these are sometimes one and the same, this does not always seem to be true. Working through these distinctions will be the subject of future work.

Part I  Rational Choice and Altruism

1  Genuine Altruism: What is it, and Does it Exist?

1.1 Introduction

Critics of rational choice theory have frequently argued that its assumptions about self-interest, or egoism, are flawed. These critics claim that we at least sometimes act to benefit others. Contemporary writers on altruism in a variety of empirical fields, such as economics, experimental game-theory, and biology, define altruism in terms of behavior. Philosophical and psychological discourse on the topic of altruism, however, has been largely concerned with motivational altruism, and in particular, with genuine motivational altruism.\(^1\) The standard way of defining genuine altruism is to contrast it with reductive altruism: genuine motivational altruism consists in having an ultimate desire or preference for the welfare of another—a desire or preference that cannot be reduced to other desires that one has for oneself.\(^2\) In particular, there has been an emphasis on distinguishing genuine altruism from hedonism, which is a form of egoism. Hedonism is the view that all of our motivations ultimately reduce to the motivation to avoid pain and seek pleasure.

---


Because hedonism seems to maintain that not wanting to feel bad when one observes another suffering is a consistently hedonistic, and therefore egoistic, motivation for action, genuine altruism has been defined in a way that omits any essential reference to feelings.

Prima facie, the standard philosophical definition of genuine altruism makes sense, for it seems we would want to reserve the ascription of altruism only for motivations that are wholly distinct from egoism. However, in this part of my dissertation I will argue that the customary way of defining altruistic motivations is unnecessarily demanding, and more importantly, misguided. The arguments I offer support a different description of genuine altruism: one in which feelings play a central role. In particular, I will argue that the emphasis on separating altruism from hedonism misses the importance of feelings to any conception of altruism. Once we include feelings in our description of altruism, the strong conceptual distinction that is said to exist between genuine altruism and hedonism collapses. Rather than identifying the specific feelings that are involved with altruism, my objective is to argue that feelings for others are essential to our understanding of genuineness when it comes to altruistic motivation. In Chapter 2, I will argue that these feelings cannot be adequately represented by, or cannot be reduced to, preferences and desires for a variety of reasons, but that they are nonetheless a kind of motivation. I will discuss there how my analysis of the central importance of feelings to altruistic motivation has broad implications for rational choice theory and its application to many forms of social and political behavior.
When discussing the possibility of altruism in the context of rational choice theory, a natural place to start is with evolutionary arguments. Rational choice theory equates what is rational with what promotes one’s welfare, and welfare is typically defined in terms of the attainment of material resources, such as money, food, shelter, and other external goods. Several authors have attempted to sever the connection between rationality and welfare by arguing that it is sometimes rational to be altruistic, and in particular, that it is evolutionarily or materially advantageous to be altruistic. For this reason, across the disciplines many discussions of egoism and altruism take place within the framework of evolutionary arguments, whether in terms of game-theory, social-psychology, or evolutionary biology: demonstrating the plausibility of the emergence or evolution of altruism in natural settings would serve a long way to show that altruism can be beneficial to the agent in some broad sense. Perhaps the most notable, systematic, and recent example of this effort is the argument advanced by Sober and Wilson in their important book Unto Others: the Evolution and Psychology of Unselfish Behavior. ³

A persistent problem for the existence of genuine altruism as defined by the standard view has been that critics can always seemingly reduce an altruistic motivation to so-called egoistic motivations— to prudential motivational explanations such as “she’s kind to them because otherwise she would not have relationships with them” or to hedonistic motivational explanations such as “he gives to charity because it feels good to give to charity” and “he feeds his children because it hurts to see one’s children go hungry”. Since motivations are opaque to researchers and since

egoism is a robust descriptive thesis about many of our biologically selected motivations, proponents of egoism argue that there simply is no need to posit genuine altruism. Researchers who take a neutral position on the question have noted, as Daniel Dennett does, that “genuine, or pure, altruism is an elusive concept, an ideal that always seems to evaporate just when you get in position to reach out to grab it. It isn’t clear what would count as genuine altruism, and paradox hovers constantly nearby.”

In the second half of their book, however, Sober and Wilson (S&W) maintain that it is plausible that not only altruistic motivation but also genuine altruistic motivation has evolved in cognitively sophisticated animals such as humans.

My arguments concerning the nature of altruism proceed as a response to the S&W book for several reasons, although I discuss where other works on altruism overlap with theirs. First, S&W adopt the standard view of the nature of genuine altruism, the view held by most philosophers, but attempt to give a novel argument to show that altruism in this sense exists. Second, their argument is an evolutionary one that is informed by the literature in game-theory, anthropology, and psychology, and has much in common with rational choice theory. I will argue, however, that their argument has a major shortcoming that points to a more general problem that evolutionary arguments for altruism confront when they adopt the standard view of altruism. Finally, S&W’s discussion spends more time on distinguishing instrumental from ultimate desires—an important feature of the standard view—than do other

---

accounts. The general goal of my arguments, however, is to discredit the egoism thesis central to rational choice theory. By showing how the standard philosophical view of altruism—even when it takes great pain to distinguish between genuine and reductive altruism and shares features in common with rational choice theory—misses something important about the nature of altruism, I hope to demonstrate that genuine altruism is not an elusive and singular phenomenon but a potentially significant range of motivations behind many of our actions. In order to do this, I must not only offer at least preliminary arguments towards a particular understanding of altruism, but must also explain why the standard view, by putting the emphasis on the wrong distinction, stacks the deck in favor of the egoism thesis of rational choice.

1.2 Standard View of Genuine Altruism

Under the standard view, and in line with common belief/desire folk psychology, what triggers or causes an altruistic preference is a belief, such as the belief that someone has been harmed. The combination of belief and desire constitutes a complete motivation to act. As will be discussed at greater length below, philosophers argue that the motivational phenomenon of altruism involves a desire. Furthermore, and important to this discussion, in order for a case of altruistic motivation to be considered ‘genuine’, it must consist in an ultimate desire. Put in its simplest formulation, an ultimate desire represents something that we want for its own sake, whereas something desired instrumentally is desired because it is a means

---

to satisfying another desire, perhaps a desire that is instrumental to still another desire and so on until we arrive at ultimate desires. Pain is one such ultimate desire because it is something we desire to avoid for its own sake—it is intrinsically undesirable. There of course may be desires that have both instrumental and ultimate dimensions. For instance, we may want to avoid pain for its own sake, and we may also desire to avoid pain in order to satisfy our desire for being productive. The desire to be productive could itself trace back to a more ultimate desire to experience a kind of pleasure, and thus the converse of pain, rendering a desire to be productive an instrumental one in the end. What many philosophers want to deny, however, is that the desire to avoid pain, or conversely the desire to seek pleasure, is the only ultimate desire for human beings. In other words, they embrace genuine altruism. Those whose works are situated in evolutionary frameworks, such as S&W, argue that a pluralistic framework of ultimate desires has been selected for in humans, and an altruistic desire is one such ultimate desire.

More generally, in the philosophical literature on altruism, and most notably in Joseph Butler, C.D. Broad, Joel Feinberg, and Sober and Wilson, definitions of altruistic motivations have been in terms of the kinds of desires they are and in particular desires that are fundamentally distinct from egoistic desires; they do not reduce to egoistic desires. Thomas Nagel’s view is distinct in that he does not define altruism in terms of desires, but rather in terms of a sense of duty. However, in Nagel’s account as in the other accounts, altruism is still defined by an ultimate as opposed to an instrumental concern. Since the differences between a sense of duty and a desire will not matter for my arguments, I will include Nagel’s account in the
standard view. In addition, I wish to briefly mention one particular work that does not fit neatly into the standard view but is salient in the contemporary literature, and this is *The Heart of Altruism* by Kristen Monroe.\(^7\) Monroe’s empirical work is noteworthy for its distinct interview approach for understanding altruism. However, there are several concerns with the way in which she characterizes altruism.\(^8\) First, she identifies altruism with a perspective rather than merely a motivation; and this seems too cognitively demanding to be an explanation for much altruistic behavior. Second, this perspective is one in which the altruist understands the “common humanity” of the potential victim; and this collapses meaningful distinctions between altruism and abiding by normative principles. Finally, the perspective is an impartial one; and this fails to allow that many of our altruistic motivations and behaviors are about loved ones and those whom we know personally, whether our families, cultures, or communities. The net result of Monroe’s conception of altruism, I believe, is to raise the bar for what counts as altruism to be too high. Altruism becomes a rare and demanding capacity that is limited to only a few unique individuals, and does not get the chance to compete as a possible widespread explanation for many of our actions.

Returning to the standard view, genuine altruism involves an ultimate desire that is other-directed and egoism involves an ultimate desire that is directed towards the self.\(^9\) But of course we need to say more about both of these desires. Taking

\(^8\) See especially chapters 10 and 11.
\(^9\) S&W want to treat certain desires, such as Jack’s desire that a resource is split between himself and Mike, as a separate category from either egoism or altruism. They label desires that are both self and other directed *relational* desires.
altruism first, I could have a desire that is other-directed, but is also malicious. For instance, I might have the desire that you lose all of your assets in a market crash. Even though this is a desire directed towards someone other than myself, we would not want to call this desire altruistic. Only desires to benefit others in some way can rightly be called altruistic. Now taking egoism, we may not want to call, for instance, a desire that oneself be tortured for its own sake an egoistic desire. It is more appropriate to say that desires that aim at the benefit of the self in some way are egoistic ones. Thus, altruism consists in an ultimate desire for the improved welfare of another, and egoism in an ultimate desire for the improved welfare of the self.  

Egoism can be divided still into further categories by giving content to the notion of welfare. For instance, welfare can consist either of any kind of external good, such as material goods only or material goods combined with public approval, distinction, or fame, or welfare can consist of internal rewards, and namely pleasure and the avoidance of pain.  

Since arguments for the first category can be more easily refuted by citing examples of an individual taking some action where there is no material or recognition returns to be had, the strongest position for an egoist is hedonism. Again, hedonism is the view that whatever action an agent performs, the ultimate motivation of the action is to avoid pain or attain pleasure. Taking hedonism as the strongest form of egoism, then, I will treat hedonism and egoism as equivalent

---

10 Egoism is silent on the question of whether individuals are motivated by short- or long- term interests or by both. Likewise, altruism is silent on the nature of the benefits to another—whether what constitutes something being beneficial can be construed as whatever I, as someone who has desires for your welfare, perceives as beneficial to you; or whether what is beneficial must be decided by the object of my desires.

except for certain occasions when it is relevant to distinguish them. For the bulk of the discussion, we are left with the definition of egoistic motivation as an ultimate desire for pleasurable feelings and altruistic motivation as an ultimate desire for the improved welfare of others. As a general descriptive thesis, ‘egoism’ asserts that the only ultimate desires we have are the desires for pleasurable feelings. The egoism thesis does not maintain that we never have desires to help others—it is obvious that we do, at least sometimes, have desires for the improved welfare of others. What egoism specifically denies is that these desires for others’ welfare are ultimate. While we can have instrumental desires for the welfare of others, the only ultimate desires are to maximize pleasure and minimize pain. The thesis of ‘altruism’ however does not maintain that the only ultimate desires we have are the desires for the welfare of others—that would be implausible and easily refutable. Rather, altruism as a descriptive thesis maintains that at least some of our ultimate desires are for the welfare of others. Put this way, it is a relatively moderate thesis that does not purport to account for all of our motivations. It does not maintain that we are all altruists all the time or all of us are altruists some of the time or even that some of us are altruists some of the time. It merely refutes the strong thesis of egoism which holds that there are never any cases of altruistic ultimate desires. Indeed, Sober and Wilson’s evolutionary arguments for the existence of genuine altruism only attempt to show that some altruistic ultimate desires have evolved. I turn to this argument next.
1.3 Evolutionary Arguments

In the first part of their book, S&W provide a strong case for the evolution of biological altruism. Biological or behavioral altruism is defined as occurring when an individual organism incurs a cost in the process of benefiting another. As understood in the biological sense, altruism merely captures a type of behavior, namely behavior that benefits another at an expense to oneself, and is silent on whether or what kind of motivations are operating. However, S&W reason that some internal mechanism must be responsible for getting the organism to behave in evolutionarily advantageous ways. After arguing for the selection of biological altruism by competition between groups, they argue that for organisms with minds, such as humans, motivations, and namely desires, are what serve to achieve altruistic behavior. What triggers or causes a desire is a belief, such as the belief that someone has been harmed. As part of the standard view, S&W maintain that the combination of belief and desire constitutes a complete motivation to act.

So what is left to determine is the nature of these desires since different internal mechanisms can be functionally equivalent in terms of causing the same kind of behavior. Indeed, the debate about genuine altruism take this problem to be central—altruistic behavior does not yet tell us whether one is a genuine altruist or not. Even if arguments and evidence can be mounted to support biological altruism through group-selection, as S&W have done, it is still a far cry from demonstrating the existence of motivational altruism. One can perform an altruistic behavior of sacrificing her own fitness for the sake of (unrelated) others and, at the same time, have done so because she experienced psychological pain from having observed the
others in harm’s way. The latter motivation would describe, according to the standard view, an egoist. Thus, one can be both a biological altruist and a motivational egoist. Furthermore, and most importantly, arguments advanced for the existence of motivational altruism do not depend on the success of arguments made on behalf of group-selection, which is still a contentious view in biology, or even on the existence of biological altruism. Biological and motivational altruism are two distinct phenomena and the domain of interest to philosophers and others interested in the rational choice debate is motivational altruism.

Historically, egoism has enjoyed a privileged status as the default assumption about motivations in the social sciences. One of the reasons for this seems to be that the social sciences have looked to the sciences, and especially biology, to inform their methodology, and egoism has been more widely accepted than views that also include altruism as an explanation of behavior in biology. When it comes to motivations, defenders of egoism argue that positing a single ultimate motivation is more parsimonious. Altruistic motivations are viewed as simply being unnecessary clutter in explaining our reasons for acting. This argument from parsimony is something like the following: pain and pleasure do such a good job of regulating our behavior because pain and pleasure are strongly correlated with harm and survival, respectively, so there simply is no need to posit any other fundamental motive than that of avoiding pain and seeking pleasure. What often happens when one is attempting to explain an action in terms of an altruistic motivation is that defenders of egoism are quick to respond with a plausible-sounding hedonistic explanation for the
action, and they claim they have parsimony on their side; this ‘hedonistic trap’ is
difficult for the defender of altruism to climb out of.

One of the reasons for the seeming ability of egoism to explain so much is its
flexibility—egoism does not tell us exactly what our motives are but rather tells us
what kinds of motives to look for, namely self-directed ones.\textsuperscript{12} So when one egoistic
explanation fails, another one can be readily offered in its place. But there is an even
more important, albeit related, source of flexibility for egoism, and for hedonism in
particular. Pain and pleasure are very vague and general psychological concepts.
Consider how broadly we think of pain. Painful feelings can be anything from simple
sensations to more complicated emotions, from an aversive sensation in one’s ear
from a loud sound to sophisticated emotions such as guilt. Pain can range from
‘passive’ sensations and feelings such as fatigue or shame to more ‘aggressive’ ones
such as hunger or disgust. It can be experienced as focused on another, as in
sympathy, or experienced as focused on the self, as in personal distress, and pain can
be coolly felt in the form of regret or can be passionately felt in the form of
resentment. Because of the vagueness and generality of the concepts of pain and
pleasure, hedonism, which holds that our single ultimate desire is to avoid pain and
seek pleasure, is also vague and general. The content of this ultimate desire is not a
single thing but includes, in the case of pain, anything we do not like, and in the case
of pleasure, anything we like.

Hedonism’s flexibility makes it far too easy for an egoist to insist that an
action was performed for hedonistic motivations whenever the actor cites any kind of

\textsuperscript{12} S&W, pp.288-289
feeling, emotion or sensation. It is not surprising therefore that philosophical
treatments of altruism have focused on carving strong boundaries between altruism
and hedonism in particular. Unfortunately, attempts to avoid falling into the
hedonistic trap have unnecessarily and to its detriment, I will argue, divorced feelings
from altruism.

Whatever we consider to be the greatest source of hedonism’s, and more
generally, of egoism’s flexibility, that it is flexible is part of the reason that egoists
can claim that it explains all of our behavior without having to posit any other
motives. Its flexibility is part and parcel of its so-called parsimony. But S&W
believe that the argument from parsimony is not a good one. They maintain that
merely counting the number of ultimate motives is insufficient for determining the
parsimony of a system, and furthermore, they remind us that parsimony is supposed
to be an all-else-equal tool of deciding between competing theories. If there are other
strong reasons, other than appealing to the standard of parsimony that is, for favoring
a theory that posits pluralistic motivations then these other reasons should be
considered and may end up tipping the scale in favor of altruism.

S&W argue that there are three main criteria other than parsimony by which
we should weigh competing theories about the evolution of certain strategies: (1)
reliability; strategies that stably contribute to fitness are more likely to have evolved,
(2) availability; only traits that are available can be subject to selection pressures, and
(3) efficiency; a trait which employs less of the organism’s energy resources has a
competitive advantage. S&W believe that the sum effect of these considerations is to

---

13 S&W, pp. 291-295
tip the scale in favor of altruism. As mentioned above, motivations can be described as the internal mechanisms for causing a human being to behave in ways that are selected for through natural selection. After arguing that selection has favored some helping behavior towards others, S&W suggest that we need to explore which motivations, and in particular desires, would have evolved in order to bring about helping behavior. They describe this inquiry as determining the solution to the design problem of altruistic behavior. I only wish to examine one of their standards—reliability—because it is this standard that they argue determines the difference between the success of altruism over egoism in explaining motivations for a particular sort of behavior. Taking this standard of reliability, they argue that a pluralistic framework that includes altruism as at least one type of ultimate desire will be more reliable in bringing about altruistic behavior than a monistic framework consisting of egoistic motivations alone.

To begin their argument about reliability, S&W make the very plausible assumption that direct internal pathways will more reliably bring about a fitness-enhancing behavior than indirect internal pathways, and they illustrate this point with an example about bacteria. Marine bacteria must avoid oxygen in order to survive. There are two kinds of mechanisms that could have evolved to allow a given bacterium to detect the presence of oxygen. One would be a straightforward and direct “oxygen-detector”. The other type would be a detector of some variable that is correlated with oxygen, such as elevation in water. They call this indirect mechanism a “toxygen-detector”. If toxygen and oxygen are perfectly correlated, then bacterium

\(^{14}\) pp. 297-327
will do just as well with either a toxygen- or an oxygen- detector. In fact, if the oxygen-detector is somehow inaccurate or flawed at detecting oxygen, then the toxygen-detector would be a more reliable guide for avoiding oxygen. But if the oxygen-detector performs at least as accurately as the toxygen-detector plus, and this is the important point, toxygen and oxygen are not perfectly correlated, then the oxygen-detector is a more reliable pathway for avoiding oxygen. Given the same level of performance of each mechanism, the direct detector (D) will be more reliable than the indirect detector (I) in producing oxygen-avoiding behavior. S&W call this conclusion the D/I Principle.

Turning to altruistic behavior, S&W assess which of two motivational pathways—an egoistic ultimate desire or altruistic ultimate desire—is more reliable in bringing about altruistic behavior. They argue that desires are a sort of instruction that tell us what actions to take given what beliefs we have. They demonstrate the contrast between altruistic ultimate desires and egoistic ultimate desires by representing the different motivational pathways involved with bringing about a case of altruistic behavior that is very widespread: parents helping their children. A parent who only has egoistic desires will take any action only if she believes that the action will maximize pleasure or minimize pain. A parent who only has altruistic desires will take an action only if she believes the action will best serve her child’s welfare. It is (plausibly) assumed that taking care of one’s children, and not merely producing them, contributes to the reproductive success of the individual. If a child perishes before he himself has the opportunity to reproduce, this limits the parent’s long-term fitness. It is important to again stress the difference between motivational and
biological altruism. Because a parent (or rather her genetic material) receives benefits from taking care of her child, it might seem as if such actions should not really count as genuine altruism. But this would be to confuse the biological with the motivational sense of altruism. Whatever the biological benefits of some behavior, what S&W, other philosophers, and indeed commentators on altruism in most disciplines aim to determine is the nature of the motivation behind those actions. Again, genuine altruism in the motivational sense consists of an ultimate altruistic desire and does not depend on whether there are benefits to the individual. The only stipulation is that the individual is not being ultimately motivated by these benefits.

Analogous to the marine bacteria example, in the case of a parent helping her child we need to determine which motivation, an egoistic one or an altruistic one, will perform more accurately in producing the fitness-enhancing behavior. S&W suggest that it is possible for parents to have flawed beliefs about what will best serve their children’s welfare, even though they do not think it is the norm. But we could also grant that the same can be said about the beliefs of the egoist. People may have flawed beliefs concerning what will maximize their pleasure or minimize their pain. In the absence of any evidence indicating which of the two kinds of beliefs are more likely to be flawed, for now we should treat the two motivational pathways as having the same level of performance.

What is left to determine is whether one of the motivational pathways is more direct than the other. Reproducing their diagram, the two different motivational paths look like the following, with the altruistic parent (A) represented in the first column, and the egoistic parent (E) in the second:
When the A parent acquires the belief that her child needs help, she formulates the desire for her child’s welfare. When the E parent acquires the belief that her child needs help, she first feels bad and as a response to this bad feeling, she formulates the desire for her child’s welfare. S&W argue that because E’s action to help her child is mediated through her feelings, there is an extra step so to speak involved when E helps her child than when A helps her child. They conclude that it is clear that an altruistic desire to improve your child’s welfare is a more direct mechanism for helping behavior than is an egoistic desire.

To relate this point back to S&W’s marine bacteria example, we can put this point a slightly different way. The object of A’s desire is her child’s welfare, so she can be said to have a “child-welfare detector”, whereas the object of E’s desire is to minimize pain/maximize pleasure, so she can be said to have a “pain-detector”. Just as toxygen must be correlated with oxygen in order for a toxygen-detector to be successful in avoiding oxygen, E’s painful feelings must correlate with her child’s welfare in order for her pain-detector to be successful in improving her child’s welfare. Recall that S&W’s D/I Principle states that given the same level of accuracy of two detection mechanisms, the direct one will be more reliable than the indirect
one in bringing about the fitness-enhancing behavior if, and again this is the important qualification, what the indirect detection mechanism detects is less than perfectly correlated with what the direct detection mechanism detects. In other words, we now need to find out whether E’s pain/pleasure is less than perfectly correlated with her child’s welfare.

S&W think there are two considerations that bear on the question of whether pain and pleasure are perfectly correlated with welfare. The first point is that some patients who have undergone a bodily injury have reported that they did not initially, and perhaps even some time afterwards, feel pain. The second, and opposite, point is that some individuals report the experience of pain in the absence of bodily harm. Nevertheless “pain is an extremely useful, but imperfectly reliable, indicator of bodily injury” when we are talking about someone’s own body. It is a different story, however, when it comes to someone detecting the pain in the body of her child. In the sort of example being discussed, E’s pain must be correlated to her child’s welfare and not her own. S&W assert that the only way this can happen is by way of beliefs. When someone’s own body is injured, she feels pain directly without having to form any beliefs about whether she was injured, and so pain-detection is a direct mechanism for producing behavior that improves her own welfare. However, when someone’s child is injured, she feels pain only after she has obtained the belief that her child is injured. So E’s pain will be even less perfectly correlated with her child’s welfare, because it is mediated by a belief, than it is with her own welfare. The conclusion that S&W draw is that since A’s motivational pathway is a more direct

\[15\] S&W, p.316
mechanism for helping her child—it is not mediated via feelings—and since E’s pain is not perfectly correlated with her child’s welfare, A’s motivational pathway is more reliable than E’s motivational pathway. S&W take this to mean that as an evolutionary strategy to solve the design problem of caring for one’s children, altruistic ultimate desires are more reliable than egoistic ultimate desires. Thus, their general conclusion is that there are strong reasons to believe that genuine altruism has evolved and exists.

1.4 Problems with Evolutionary Arguments for Genuine Altruism

Despite the seeming plausibility of S&W’s arguments, and the optimistic conclusion they draw, there are several problems with their arguments. The first point I want to make concerns the role of beliefs in the egoism pathway. S&W argue that desires operate in the example above only through a belief that is relevant to the object of desire. In order for E to feel bad when her child needs help, she must first form the belief that her child needs help. Because of the role of this belief, a case of a parent responding to the harm to her child is very different from a case of someone responding to her own bodily harm. When I injure my leg, I do not need to first form a belief that I have injured my leg in order to feel pain. I feel pain as a direct result of what is happening to my body. E’s painful feelings about her child, S&W argue, are not nearly as direct.

Contrary to what S&W maintain, however, it seems that there are at least some circumstances when E’s painful feelings about what is happening to her child could occur in the absence of forming any belief. Consider when someone you know,
or even someone you do not know, is performing in some way—perhaps dancing on stage or delivering a presentation. If this person looks like he is about to trip and fall in the middle of his performance but then stabilizes, you might have a tingling sensation in your gut, and a quickening of your heartbeat some time after his near-fall, despite the fact that you do not have the belief that he fell or will fall. There are better examples. Think about a movie where a character is being tortured by a malicious killer. Even if you consciously think to yourself that what appears to be happening is not in fact happening, you might actually experience pain in the parts of your body that correspond to where the character is being tortured. It would seem odd if I were to insist that you must be having the belief that this person on your screen is actually being tortured or has been tortured.

One might think that the dissimilarity between pain experienced in our bodies due to our own bodily harm and the possible pain experienced in our bodies in response to the bodily harm in others is still too strong. After all, our immediacy with our own bodies’ pain is what grants pain its superior epistemic status—we are wired to experience the pain of our own bodies for the purpose of informing us of bodily harm. But if we broaden our notion of what constitutes pain, the difference seems to go away. As S&W themselves acknowledge, egoism does not hold that the category of pain just applies to our physical sensations. Negative emotions, feelings, and perhaps even moods, if they are sufficiently aversive, count as pains. As I mentioned above, this broadening of the notion of pain is what makes egoism, and in particular hedonism, such a flexible view. But once we include emotions and feelings, it seems more obvious that there are occasions in which we experience pain in response to the
suffering of others without having to first form beliefs. How many times have you cried by watching sad movies or by reading sad novels which are fiction? You might insist that you are forming some belief, perhaps that the plight of the character resonates because you are aware that others are experiencing that plight in the real world. Or perhaps, we are really being duped into believing that these actors on the screen are suffering, if only for a moment.

We should be skeptical of these responses, but even if one holds their ground, there is evidence of ‘emotion contagion’ even when the circumstances do not involve fictional characters. For instance, social-psychologists have studied people’s facial expressions as participants observe the expressions of others and found that there are different response patterns, as measured by electromyographic (EMG) analysis, when participants observed happy versus angry faces.\(^{16}\) And the connection between facial expressions and the experience of emotions appears to be direct and specific. Paul Ekman and his colleagues had participants either recall and try to relive experiences of a certain emotional kind, such as anger or sadness, or, arrange their facial muscles in ways that are characteristically associated with these emotions. The researchers found that both groups of participants experienced the autonomic nervous system (ANS) responses that are typically associated with the particular emotion.\(^{17}\) Emotion contagion frequently appears to occur automatically, without the mediation of belief or other conscious processes. Mark Davis argues that the synchrony of a multitude of facial, vocal and postural responses is too fast and too extensive to always be

mediated through conscious processes. He argues that it probably involves multiple-levels and areas of the brain simultaneously. In one study, participants mimicked and synchronized the postures and movements of others within 21 milliseconds. That’s a pretty speedy rate when you consider that in practice sessions it took Muhammed Ali at least 190 milliseconds to detect a light and 40 more milliseconds to respond by throwing a punch—230 milliseconds for something that we might reasonably put in the category of cases of a person reacting to threats to his own bodily injury, and not to that of others.

Maybe the best example of emotion contagion happens within the first few days of our lives. Consider how frequent it is for newborn babies to cry when other babies are crying. Whether crying-contagion is a form of empathy or sympathy or other complex emotions, or even any emotion at all, is a matter of some controversy. But it seems that babies are experiencing some distress, however self-focused, when they cry in response to other babies crying. Distress is an aversive feeling, and the hedonist has no problem accepting that babies feel distressed simply by perceiving the distress of others without first forming a belief.

If babies experience aversive feelings simply by registering the distress of other, unrelated babies, it seems just as plausible that a baby’s mother, or father, is capable of experiencing aversive feelings when her, or his, own baby cries simply by

---

registering her babies’ distress. Let us take a look at how the egoistic (E) parent’s motivational pathway might be different if her feelings did not have to be mediated by beliefs.

**Figure 2**

Child needs help

↓

↓

A’s beliefs about Child

E’s feelings about Child

↓

↓

A’s **Desire** for Child’s Welfare

E’s **Desire** to Minimize Pain

↓

↓

A Helps Child

E Helps Child

I have presented E’s feelings as occurring directly in response to observing that her child needs help, and E’s desire is mediated through her feelings. Furthermore, precisely because the view of altruism under discussion is one that does not require feelings or sensations, A’s desire is not presented as being mediated by feelings or sensations, but by a belief. If E’s motivational pathway operates in this way, then notice that it takes E just as many steps as it takes A to help their respective children. In other words, A is not a more direct solution to the problem of helping behavior in parents, and without further argument, A is not more reliable than E. None of what I have just discussed implies that all of E’s pain for her child occurs in the absence of beliefs. If E is in a different country than her child and hears from a third party that her child is injured, E’s aversive feelings might very well only occur by first forming a belief about her child. Nevertheless it seems that some, possibly
many, of E’s feelings about her child’s welfare could occur without the prior formulation of a belief.

In the first part of their book, S&W argue that altruistic behavior or biological altruism can be selected for because of its advantageousness for one’s kin or group. Likewise, having the capacity to immediately and directly feel pain in response to perceiving that others are in pain could plausibly be selected for because of the same reason—because it is advantageous for one’s kin or group. It is efficient for an organism to feel pain when her hand is burning and withdraw her hand without her first having to formulate a belief that her hand is burning. Just as it would be inefficient for her withdrawing of her hand to be keyed to her beliefs, it would also be inefficient, from the standpoint of caring for one’s children, for a parent to remove her child’s hand from a fire only after forming a belief. My aim here is not to argue one way or the other that pain in response to another’s pain has been selected for or that there is an evolutionary advantage to such a response. My aim is only to point out an alternative understanding of pain response that would not be ruled out, and could be supported, by the type of evolutionary argument that S&W advance. More generally, it is to convey a problem with relying on evolutionary arguments to support the existence of altruism. Aside from the possibility of an evolutionary argument for the existence of some trait being a “just-so” story, there is a further problem more specific to the phenomenon of genuine altruism. I will turn to this more serious problem with S&W’s argument next.

Recall S&W’s D/I Principle as applied to altruistic and egoistic motives. It states that given the same level of accuracy of the two motivational pathways, the
direct one will be more reliable than the indirect one in causing parents to take care of their children if there is a less than perfect correlation between what the indirect motivation is keyed into and a child’s welfare. A desire for the improvement of another’s welfare is a direct internal mechanism for causing helping behavior and a desire to avoid pain or seek pleasure is an indirect mechanism. Assuming S&W are correct in arguing that feeling pain in response to another’s pain only comes after belief-formation, then the reason for the indirectness in the egoism pathway is the additional step it involves. Altruistic desires are therefore more reliable in promoting other-directed behavior than egoistic desires.

The problem with S&W’s argument, however, is that the satisfaction of the D/I Principle only requires an altruistic desire. It is silent on the nature of this desire. In particular, it does not tell us whether the desire has to be an instrumental or an ultimate altruistic desire. In fact, if A’s desire for the improvement of her child’s welfare were an instrumental desire, this would satisfy the D/I criteria. It could be a desire that serves another more ultimate desire, however near or remote, and because the role this ultimate desire plays in A’s motivation is to bring about the existence of other desires, it operates at a step removed. It is not directly tapped into, so to speak, when A’s child needs help. Instead A’s desire to help her child is what gets activated. This is not to suggest that A’s ultimate desire would be an unconscious one—although, as I will discuss below, S&W think even this is perfectly possible and should not be ruled out. It only suggests that A’s ultimate desire would not be experienced at that time, whereas E would experience her ultimate desire to avoid pain—this is what would account for the indirectness of E’s motivational pathway.
But, and this is the important point, even though A’s desire to help her child is more direct than E’s desire to avoid pain, A’s desire would still only be an instrumental one. Presumably, A’s desire would be relatively more ultimate than E’s desire, since, again, E would also experience the desire to avoid pain. However, on the view that genuine altruism involves an ultimate altruistic desire, the fact that A’s desire to protect her child would be relatively more ultimate than E’s does not yet tell us whether A is a genuine altruist or not.

A might have desires for many things that we would want to call altruistic, each with a different object, such as to promote peace, to read literature to disadvantaged youth, to look out for her friends, and to promote the wellbeing of her child. But as long as these desires only exist as means to bring about the satisfaction of other, more fundamental, desires, they are all instrumental desires. In fact, she could, ultimately, have only the single desire to minimize pain and maximize pleasure.

This observation poses a serious problem for S&W’s effort to establish the evolutionary reliability of genuine altruism since their definition of genuine altruism entails an ultimate, and not an instrumental, desire. If an ultimate desire to minimize pain/maximize pleasure is what gives rise to A’s desire to protect her children, then A is not an altruist and S&W have not shown us why it is plausible to think that genuine altruism has evolved. More generally, the difficulties facing any evolutionary argument for the existence of genuine altruism should now be clear. Despite being convincing arguments for the emergence of altruistic motivation, evolutionary arguments such as S&W’s could not hope to show us why this altruistic motivation
would come in the form of ultimate and not instrumental desires, since, as I have demonstrated with the alternative interpretation of A’s desires, the two are functionally equivalent. Instrumental desires serve as reliably as ultimate ones in producing altruistic behaviors when we, plausibly, maintain that ultimate desires can be background ones that are not always experienced at a given time.

1.5 Conclusion

So where does this leave us? S&W offer the most convincing argument to date in support of the notion that genuine altruism has evolved, and more generally, that it exists. Since their arguments fall short, does this mean we are doomed to believe that there are no good arguments demonstrating the existence of genuine altruism and so we must believe that we are all, ultimately, egoists? I do not think so. Arguments such as S&W’s fail to be convincing because they employ the standard view of genuine altruism, which forces us to parse out ultimate from instrumental desires. I will argue, however, that when we reflect a little more on the nature of altruism, the view that the genuineness of altruism depends on the level of desire—or whether a desire is ultimate—is not a compelling view of altruism. In other words, it is not the case that in order to argue for the existence of genuine altruism we will have to accept a less accurate notion of what we mean by altruism.

It seems the greatest strength of S&W’s argument for the evolution of motivational altruism has to do with the flexibility of a system in which altruism would be embedded. A pluralistic motivation system is one that would include both egoistic and altruistic motivations. S&W reason that pluralism with respect to
motivations is more plausible, from an evolutionary perspective, than monism. Kim Sterelny has made a similar argument about the benefits of plasticity that come from motivational pluralism.\textsuperscript{21} A pluralistic parent, one with both altruistic and egoistic desires, will more reliably help her children than a purely egoistic parent. Whereas the egoistic parent only helps her child when it minimizes pain, the pluralistic parent will help her child when it minimizes pain or when it promotes her child’s welfare. If one motivational pathway fails, there is another one to do the job. In some cases, helping her child will both minimize pain and promote her child’s welfare. So she will have two reasons to help her child. As long as the two pathways operate independently so that they do not interfere with each other, the two pathways are more reliable than either pathway working alone.\textsuperscript{22}

The evolutionary arguments for motivational pluralism in cognitively sophisticated beings such as humans are strong, but it is not necessary that the kind of pluralism that has evolved pertains to different ultimate desires. The same arguments support the evolution of pluralism with respect to different emotions or feelings, which, I will argue next, are a distinct kind of motivation from desires. Since I will argue that emotions rather than ultimate desires play a prominent role in our conception of genuine altruism, it is important to spend some time, in the following Appendix, saying what emotions are and how they differ from desires. In addition,


\textsuperscript{22} I can think of cases where the two pathways might conflict. I might feel bad punishing my child for something she has done but at the same time believe that it is promotes her welfare in the long-run. How would I choose? Maybe a purely egoistic parent would face similar conflicts. I might feel bad punishing my child but at the same time feel pleasure from the pride of being a firm parent. While comparability might be somewhat less complicated for a pure egoist, the argument from pluralism is a good one.
my analysis of the implications of my view of genuine altruism in the Appendix to Chapter 2 builds off of this discussion on the emotions. However, if the reader chooses to skip this section for now and go directly to the discussion of the philosophical arguments for the existence of genuine altruism, she may proceed to the beginning of Chapter 2.
1.6 Appendix 1: Are Feelings and Desires Separate?

In *Gut Reactions*, Jesse Prinz provides a discussion of the varying views on emotions, focusing on all dimensions important to each prominent view. In order to sort through them, I begin by organizing the kinds of theories around the two dimensions that matter most to my discussion here: 1) motivation; because the debate between genuine altruism and egoism is a debate about what motivates our behavior; and, 2) feeling; because it is this dimension that I will argue is missing from the standard view of altruism. Below is a representation of the typology of emotion views as they pertain to these two dimensions.

<table>
<thead>
<tr>
<th>No Feeling</th>
<th>No Motivation</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1: Cognitive Views</td>
<td>2: Behaviorist View</td>
</tr>
<tr>
<td>Feeling</td>
<td>3: Somatic Views</td>
<td>4: Hume</td>
</tr>
</tbody>
</table>

According to Prinz, many theories of emotions maintain that emotions have motivational force. For instance, a variety of views hold emotions to be action-tendencies. At one end of this spectrum are certain behaviorist views, represented in cell 2, where emotions are only dispositions to act in certain ways. In other words, emotions are described in the same way that many philosophers describe desires. These same behaviorist views also maintain that there is no feeling element to emotions. Some views divorce emotions from both feelings and motivations. For instance, Cell 1 contains cognitive views, which hold that emotions are judgments like beliefs, such as a belief that one has been wronged, or are judgments about judgments, such as an evaluation that the belief that one has been wronged is
justified. Both the Cognitive and the Behaviorist views lack what seems to be intuitively central to emotions, that emotions involve feelings. In one study, every group of people surveyed except one rated the feeling component of emotions to be the most important. Whether or not feelings accurately capture the essence of every emotion, it is intuitive that feelings are central to many. In any case, feelings of a certain type will end up being important for the discussion of altruism, and therefore so will views of emotions that give a central role to feelings. There are two kinds of views that do this.

At the intersection of both the motivation and the feeling dimension is something like Hume’s view, where emotions are feelings that contain desires, and desires according to Prinz are themselves a subset of feelings that impel us to act. Prinz writes that “fear may contain the felt desire to flee." A desire to flee is, for Hume, an action tendency.” A great deal has been written about the nature of desire in Hume. That all desires for Hume have a phenomenological feel is a matter of dispute. Michael Smith interprets Hume as maintaining that only some desires are experienced as feelings. He denies that Hume’s desires are only or even always feelings. More broadly, even aside from his reading of Hume, Smith makes a compelling case for the implausibility of viewing all desires as having a phenomenological feel. He asks us to consider a long-term desire, for instance, the long-term desire to be a doctor. In moments of reaffirmation of this desire, following

---

23 Jaak Panksepp, 2000. Subjects were asked to rank possible emotion components such as action-tendencies, feelings, thoughts and bodily changes. The group that failed to rate feelings as most important were college students majoring in philosophy. This group rated thoughts to be most important.
24 Prinz, p. 11
periods of doubt, this desire to practice medicine might express itself with a kind of force and perhaps even a feeling. But even in the absence of this feeling, during times in which the desire is not being challenged or reassessed, we would not want to say that this desire simply goes away. Furthermore, when someone is sleeping we would not say that their desires simply cease. However, as Smith points out, feelings may turn out to be central to certain desires, such as a desire to eat. In the dispositional account of desires presented above, these desires would include a disposition to feel a certain way in addition to a disposition to act in particular ways.

While desires do not always involve feelings for Hume, according to Smith, Prinz argues that feelings are central to Hume’s account of the emotions. But they may contain more than this. Emotions for Hume also seem to contain cognitions insofar as thoughts are used to distinguish one emotion from another. If this is the case, then Hume’s view is also partly a cognitive theory. But, as Prinz has convincingly suggested, we should not accept cognitive theories of emotions. A moment’s reflection seems to dispel the notion that cognitions are defining aspects of emotions. He offers the following example to support his point. “A friend has convinced you to go on a hike. You are not much for the outdoors because you suffer from a terrible snake phobia. Reluctantly, you go along. As you set out, all concerns quickly drift away. You are having a good time. But then, out of the corner of your eye, you glimpse a long coiled object nestled in the shadows. You freeze, your heart pounds, your breathing becomes strained, your eyes widen.”

Prinz’s point is that despite not thinking about snakes, not having any cognitions related to snakes

---

Prinz, p. 21
whatsoever, you can imagine yourself becoming fearful the moment you spot something that looks like a snake; even, perhaps, continue to be afraid for a few moments after you acquire the belief that the object is not a snake. Prinz does not argue that cognitions never cause emotions—he certainly believes they can—he just maintains that cognition is not a necessary cause of emotion, and that cognitions are not a part of emotions. This brings me to the final type of view, a view that Prinz has called the “somatic feeling theory”.

Beginning with William James and Karl Lange, somatic theorists have maintained that emotions respond to, register, or track changes in the body. Each theorist has his own take on the nature of this response. For instance, for Antonio Damasio, feelings are representations of bodily changes. If an organism is in the presence of a predator, a harmful situation, its body will respond with an increased heart rate, blood-flow and skin conductance, and the organism will begin to flee. These bodily changes help organisms navigate the world by responding successfully to stimuli. But then what work do feelings do? Feelings enter into the picture by allowing the organism to be aware of these bodily changes and by giving the organism an incentive to act. Feelings also help an organism use combined responses to past situations in order to form novel responses for situations that have not happened yet.  

Although Damasio argues that feelings are important for their role in representing bodily changes, he maintains that emotions need not involve conscious

---

27 For Damasio, emotional feelings contribute to another step in the development of sophisticated cognition: the feeling of knowing that we have feelings. What is the feeling of knowing? This is core consciousness. Damasio argues that consciousness evolved because it allowed organisms to feel feelings, or, in other words, to be aware of their emotions.
experiences. The brain can perceive bodily changes without causing feelings. Emotions simply are the bodily changes that respond to stimuli in the environment, and the feelings they give rise to are separate. To echo an earlier comment, what we choose to include under the specific label of emotions is not terribly important for my purposes. What I am interested in is this theory and others where feelings are importantly linked to emotions, whether causally or constitutively. It happens to be the case, however, that for Prinz’s theory of emotions, feelings also fall under the label of emotions and are exhaustive of emotions. For this reason and for more substantive reasons, his is the theory I favor. However I will offer certain amendments to his theory.

Prinz argues that emotions are feelings that appraise certain situations, and they do this by tracking bodily states. It’s more complicated than Damasio’s view where emotions simply represent changes in the body, and so it deserves to be fleshed out. Prinz argues that for emotions to represent something, they must be reliably caused by that thing. Above, I briefly demonstrated how Prinz successfully argues for the somatic view, so it would seem obvious that what reliably causes emotions in Prinz’s view is the body, and that what emotions represent are bodily states. However, Prinz also argues that for something to represent something else, it must have the function of representing that something else, and it seems implausible that emotions have the function of representing bodily changes.

If for instance I site a snake, my heart races, I feel afraid, and I run. Prinz argues that it would be a mistake to think that my feeling of being afraid represents my heart beat. My emotions represent something, but not, contrary to Damasio,
bodily changes. What would the survival advantage of that be? Prinz maintains that emotions have the function of representing real situations. To see this, we can isolate each, the body and the situation, to discover what it makes more sense for evolution to have settled upon. Taking the body first, if feelings were to have the function of representing bodily changes, then it is plausible that I would feel afraid, and flee, simply if my heart beats faster, without there being any real situation that ought to cause me to flee. That does not seem to make much sense. However, what if I feel afraid, and flee, in the absence of any heart-rate increase, but because I confront a situation that ought to cause me to flee? Now that would appear to be advantageous. So what sort of situation ought to cause me to flee according to Prinz? One where there is, or might be, a real danger present, like a snake. This kind of situation is, more generally, “a relation between an organism and its environment that bears on wellbeing”.  

Emotions are feelings that represent relations between an organism and its environments that bear on its wellbeing. For instance, as already indicated, fear represents danger to the organism and sadness represents a loss to the organism, and the organism registers danger and loss through bodily changes.

To return to the point of my discussion on the emotions, I was exploring the role of motivation and feelings in the emotions. Prinz’ view renders feelings essential. But what about motivation? In addition to representing organism-environment relations that bear on wellbeing, there is another component to emotions in the Prinz view—valence. Valence is the intuitive idea that every emotion is either a negative or positive emotion, either agreeable or disagreeable. Sadness, anger, and

---

28 P. 51
fear are negative emotions, while happiness and joy are positive, and certain emotions such as surprise might be mixed. But according to Prinz, emotions do not have negative or positive valence in virtue of the way they feel to us, for valence could be an unconscious aspect to emotion. While I think this is debatable, what is more relevant to my discussion on altruism is what Prinz takes to be the defining feature of valence. For Prinz, the valence of an emotion is like an inner command to either want more or less of a situation, all things considered. I have added the “all-things-considered” qualification because, as Prinz notes, there may be other considerations that outweigh the command. Echoing Damasio’s position described above, the valence aspect of emotions is more like an incentive: emotions give us an incentive to act. To sum up Prinz’s full view of emotions then, and adding the valence component, the feeling of fear would not only represent a dangerous situation, again, through registering changes in the body, but it would also represent the danger as something that is bad or negative, and serve as an inner reinforcer to withdraw from the dangerous object; to eliminate the situation.

Because it urges us to want more or less of an object or situation, valence sounds an awful lot like a motivation. Prinz is sympathetic to this. However he thinks there is a useful distinction to draw. Several emotions, he suggests, are not always action-tendencies. For instance, he argues that anger does not always motivate us to seek revenge and fear does not always impel us to flee. He notes that all emotions, due to their valence marker, provide a reason for acting, but do not always in

---

29 Prinz doesn’t deny that emotions feel unpleasant or pleasant. Anger is a disagreeable emotion. It feels unpleasant. He just denies that there is one general feeling of unpleasantness or pleasantness that characterizes all negative or positive emotions respectively.

30 p. 193
themselves impel us to act. This is the essential difference between emotion and motivation. When we are angry, the valence marker issues its command, “less of this!”, but it is a command directed toward an inner condition (an angry mental state) but it is not a command for action. Emotions facilitate, or prepare us for, action, through bodily changes. But emotions do not consist of decisions, plans, or choices to act, which have external situations and states as their object—that is the realm of motivation. Prinz does believe the link between motivation and emotions is tight: he denies that every action we take is a motivation—consider autonomic responses such as my moving my fingers while I type—and wants to save the term motivation only for responses that follow affective states, where “hedonic considerations” apply. Since emotions are affective states, a motivation will frequently result from an emotional state. But other affective states, such as hunger or fatigue, can trigger a motivation and some emotions, again such as anger, will not always incite a motivation. The two concepts can come apart.

I disagree with Prinz’s assessment here. There seem to be different senses of motivation, and thus room to view the emotions as motivations of a certain type. Furthermore, Prinz’s denial that emotions are motivations seems to stem from his conflation of at least two of these senses of motivation—motivation as force, and motivation as decision-making. More specifically, he employs different notions of “action-tendency”, which is what he takes to be central to motivation. On the one hand, Prinz seems to hold the view that action-tendency is like a force. It is difficult
to explain this notion well, but it is a familiar intuition. It is common to speak of motivation in terms of having an “urge” or “drive”, or being “impelled” or “pushed” to act. This seems to be the idea of action-tendency behind Prinz’s statement that hunger, which he defines as a motivation, “commands us to eat” and is an “impulse to eat” and his view that motivations are always triggered by affective states. On the other hand, Prinz also identifies action-tendency with terms such as decision and choice, suggesting something that is more active than a mere tendency. For it seems not only that many decisions and choices lack any kind of force of their own, but also that they are realized subsequent to a tendency. For instance, when I decide to pick up my tea cup, it would be more accurate to describe my wanting tea as a disposition and my deciding to pick up my tea cup as what follows from my disposition. It is the course of action I choose in order to bring about the satisfaction of my want. A decision to pick up my tea cup and take a sip or a choice to change the television channel seem to have a stronger kinship with behavior and action than with a disposition or tendency for action—on the assumption that tendency or disposition for action is causally prior to action.

When Prinz denies that emotions are motivations he employs this sense of motivation as decision-making. He argues that in the case of the emotion anger, for

---

31 Perhaps the difference between the two views of motivation can be brought out by noting what is implied by the passive versus the active voice. A motivation seems to be more like Prinz’s use of the term motive—a reason for action. But to say that something is motivating seems to imply that it is compelling you to act.
32 Ibid, p. 194-5
33 As Prinz notes, most of these terms—decision, want, desire, goal, hope, choice, etc.—are multiple-meaning terms. Again, it is not the exact term we employ, but the spirit of the term that seems to matter.
34 Notice that this doesn’t seem to be the same thing as my instrumental desire to pick up my teacup, which I would maintain is a motivation.
instance, our bodies are prepared for action, but “at this point in processing, no action has been selected, no strategy has been determined, no plan has been conceived. The somatic state and valence marker must be fed into a mental system that selects responses. Among the available responses is violent revenge against the source of our anger…Once that choice has been made, we can say there is action tendency at work.”\(^{35}\) If you view motivation as decision-making, then it is not surprising that emotions under Prinz’s view would not count as motivations. For in the very plausible somatic account of emotions that Prinz develops, emotions are bodily feelings that merely represent, and in a sense respond to, organism-environment relations, and this hardly seems to imply decision-making. But the decision-making view of motivation is very different from a view of motivation as an urge, force or push or pull. The two seem to be at odds with each other, for one involves a picture of an agent as passive and as being acted on, and the other pictures the agent as active.

Rather than reconciling the two notions, it seems plausible to maintain that there are different kinds of motivation. This point can be stressed by noting that neither of the two accounts seems complete on its own. I have already discussed how the decision-making account maintains a condition that seems too active for everything we would want to call motivation. The force-view does not seem completely right either. If action-tendency is a force then it seems that many of our motivation terms, and many of the ways in which we explain or make sense of someone’s actions, would not qualify as motivation. For instance, I currently have a want to continue typing, but it would be misleading to think of this want in terms of a

\(^{35}\) Ibid., p. 194
“force” that is triggered by an affective state. Perhaps this want is instrumental to a more ultimate want, such as wanting to finish my dissertation, for which the notion of force or being impelled would be appropriate as a description. Nevertheless, I think it is plausible to view my, albeit instrumental, want to type in this phenomenologically weaker way. But how would you describe my wanting to continue typing if not as a motivation? It seems we should allow for other kinds of motivations than only affectively situated ones.

The insistence that there is only one sense of motivation fails to capture our ordinary usage of a variety of motivation terms. Perhaps certain motivation is more like decision-making, and another type is more akin to the non-affectively charged desires and wants that are necessary to make sense of or to explain behavior, and still another kind feels like a force, one over which the agent seems to have little control. It seems emotions would qualify as motivations of this last type. Let us revisit Prinz’s comments regarding forces or urges. He argued that emotions do not always urge us to act, and for support he cites the occasional anger that fails to impel us to seek revenge and fear that sometimes does not urge us to flee. In fact, I would maintain that this is frequently the case with anger and fear. Most of the time I am angry, I exhale loudly and curse, and these actions are a long way from seeking revenge. But, contrary to decisions or choices, urges and forces do not seem to have specific actions as objects. They are commands to do something in response, be it kick a shoe or exclaim rage or to huff and puff. However, they need not always compel us to take a particular action, such as seeking revenge.
Furthermore, unlike urges such as hunger or fatigue, any given emotion represents a variety of different organism-environment relations, even if only subtle variations on the same kinds of theme, and so the behavior that emotions command will vary in response. To use anger as the example again, Prinz argues along with several other emotion researchers that the kind of thing anger represents is an offense to me or to those close to me. This seems correct, but offenses can come in different forms and in different degrees, and these differences appear to have an effect on the behaviors that anger commands. Hunger represents something quite concrete and unvarying, namely, undernourishment, and no matter the degree of hunger, the action command is the same: to eat. Anger, however, can represent varying kinds of organism-environment relations, with more or less offensiveness and distinct kinds of offenses—I can imagine being so angry at someone that I am compelled to avoid him. It is not surprising therefore that anger sometimes fails to urge us to seek revenge. Nonetheless, it compels us to do something.

To sum up my discussion about the relationship between emotions and desires, I conclude that emotions are feelings and motivations, but they count as motivations of only a particular kind—as a force or urge for action. While there is this motivational force to emotions, this does not entail the more robust view that emotions contain desires. Desires and emotions, or at the very least, feelings, seem to be distinct. As we will see in the next chapter, this means that the standard view of genuine altruism makes no essential reference to feelings of any kind. I will argue that this is a significant shortcoming of the view. Furthermore, once we add feelings to the picture, it becomes apparent that genuine altruism is more widespread than has
been traditionally assumed. Finally, in the second Appendix below, I will demonstrate how including feelings as motivations as part of the conception of genuine altruism may open up several promising avenues for both theoretical and empirical research.
Part I Rational Choice and Altruism

2 Ultimate or Instrumental Altruism? Who Cares

2.1 Introduction

I argued in the previous chapter that evolutionary strategies for demonstrating that genuine altruism exists are problematic, and ultimately, fail to be convincing. In this chapter, I will now turn to general philosophical arguments on the subject and show why they too fail. I will argue that the source of the failure to establish the existence of genuine altruism is the way in which genuine altruism is defined. As stated in Chapter 1, the standard philosophical conception of altruistic motivation is that it is a desire for the improved welfare of another, and genuine altruistic motivation is an ultimate or non-reducible desire. Again, what triggers or causes an altruistic desire is a belief, such as the belief that someone has been harmed. In a similar manner, we could describe many instances of desires that are triggered directly by beliefs alone. For instance, my desire to wear a raincoat is keyed directly to the belief that it is raining outside. While it might be perfectly accurate to describe the desire to wear a raincoat as not involving, and never involving, feelings, is it accurate to describe altruism in this way, as occurring in the absence of feelings, without any role for feelings? I do not think it is.

Historically, altruism has been associated with certain passions. Aristotle’s description of *filia*, usually translated as “friendship” or “love,” is the passion that describes the condition of sharing another’s pleasures and pains and he also believes that good men feel pity for those who experience undeserved suffering. Both Adam
Smith and David Hume of course recognized our tendencies for fellow-feeling and for perceiving the suffering of others through our ability to sympathize.\(^1\) Joseph Butler, who wrote extensively on the topic, identified altruism, or benevolence, with certain affections and love.\(^2\) Yet, despite some associations between altruism and certain feelings, the standard view of altruism makes no essential reference to feelings, and thus underplays their significance to altruism. In this chapter, I will argue that the standard view does not present a sufficient understanding of genuine altruistic motivation. In particular, whether a desire is ultimate may be an irrelevant question for two reasons: first, a desire might not be ultimate but might be genuine nonetheless; and, second, feelings are just as, if not more, important to the conception of genuine altruism.

### 2.2 Prior Suitability

To begin my analysis of the standard view, I want to introduce a major philosophical argument against the thesis of egoism. Some authors, including S&W, are unimpressed with this argument against egoism. But as I will suggest later in the chapter, this argument may be revitalized once we consider the role of feelings. A line of philosophers, starting with Joseph Butler and continuing with C.D. Broad, Joel Feinberg, and Thomas Nagel, argue that many apparently hedonistic desires are not egoistic upon closer analysis. They argue that the only way we can gain pleasure from

---

1. It is often argued that their notion of sympathy is much more akin to the way we use the term empathy today, as connoting a capacity for experiencing what others experience and identifying with others.
2. Butler, *Five Sermons*
certain experiences is if we already desired something else, and that desire was satisfied. For instance, Butler writes “there could not be this pleasure, were it not for that prior suitableness between the object and the passion: there could be no enjoyment or delight from one thing more than another, from eating food more than from swallowing a stone, if there were not an affection or appetite to one thing more than another.” Other philosophers reason slightly differently, but the general argument, which I call the ‘prior suitability’ argument, is the following: since someone feels good after the satisfaction of a certain desire, say the desire for food, this demonstrates that the desire existed prior to and independently of the desire to feel good. The causal chain describing the pleasure we receive when we eat food would look like the following:

Desire for Food → Eating → Pleasure

If pleasure is something that occurs as a result of the desire for food, then the pleasure did not cause the desire for food. Furthermore, it seems someone could easily maintain that her desire for food is an intrinsically-held desire, something that is desired for its own sake, in which case, such a person, at least when it comes to her desire for food, would not have a hedonistic desire.

This all seems true enough, but note that prior suitability between a desire for food and the pleasure gained from eating food does not refute hedonism. The prior suitability line of reasoning does not tell us anything about the relationship between the desire for food and the ultimate desire for pleasure. If the desire for food is

\[ ^3 \text{Butler 1726, p. 227} \]
\[ ^4 \text{S&W, p. 278} \]
caused by the ultimate desire for pleasure, then Butler’s reasoning is unconvincing as a refutation of hedonism. So we need to know more. Butler’s argument raises the following question, which I call the ‘prior cause’ question: what caused the desire for food? If this explanation appeals to internal rewards, and for most people most of the time it seems it would, then it would appear that hedonism, and hence egoism, is vindicated right? Well, actually, there is another way to interpret the internal rewards response, and perhaps even Butler’s prior suitability argument, that does not implicate hedonism. What if the answer to the prior cause question does not cite a more ultimate desire for internal rewards, or any desire at all? What if the answer cites a feeling or emotion rather than a desire?

2.3 The Central Role of Feelings

The Butlerian type of argument that attempts to rule out hedonistic explanations holds that the fact that someone feels pleasure after the satisfaction of a certain desire, say the desire for food, demonstrates that the desire existed prior to the desire to feel good. But the prior suitability argument is not sufficient to refute hedonism because we will still want to know what caused the desire for food. If the explanation refers to internal rewards, then it would appear that hedonism, and hence egoism, is vindicated.

As I just suggested above, however, the answer to the prior-cause question does not need to cite a more ultimate desire for internal rewards, or any desire at all. What if the answer cites a feeling rather than a desire? In that case, it seems we do not
need to press further about more ultimate desires. Instead, it seems our inquiry at that point ought to be one of understanding the nature of the underlying cause of the feeling. There are three separate reasons for this, and more generally, for diverging from the standard view of altruism underlying the obstacle to the prior suitability argument. I turn to these three reasons next.

2.3.1 Reason 1

The first point is a point of language, but is not only a point of language. It is also an observation that might lend support to the hypothesis that genuine altruism is widespread. Since the question of altruism is not just about the existence of any altruism (one instance of genuine altruism in the history of humanity would presumably be insufficient to convince skeptics) but also about its prevalence, this will turn out to be an important point. We frequently use the language of hedonism to describe our motivations in very general terms, masking more specific explanations that, on closer inspection, may turn out to be anything but egoism. Consider how we often describe a reason for helping someone in terms of the feelings that are stirred by seeing someone being harmed. As mentioned above, this fact gives egoists fodder for believing that most people are not genuine altruists. But inspecting the full context of a response like this, including queries for further explanations and subsequent responses, may support the view that what lurks behind this response is genuine altruism indeed.
To illustrate this point, I return to S&W’s example of a parent tending to her child’s welfare. If I have a desire to take care of my child (an altruistic desire) and one asks “why do you want to take care of your child?”, if there is no further desire to which to refer then the desire to take care of my child is ultimate, and according to S&W, it is the real deal—genuine altruism. On the other hand, if I give an explanation that cites my feelings, such as “because it feels good to take care of my child”, then the egoist could argue that this implies that I have the further, more ultimate desire to feel good and so my desire to take care of my child is merely an instrumental desire, one along the road to egoistic hedonism. However, if I were to say that I feel bad when my child is harmed and this is the reason that I take action to ensure his welfare, this is not enough for the proponent of egoism to win the debate in terms of the standard definition of altruism. For the egoist to win, he must get me to explain my feeling in terms of a prior desire. This point can become clear by focusing on two different ways of probing further into my feelings:

Question 1: “why does it matter to you that it feel good to take care of your child?”
Question 2: “why does it feel good to take care of your child?”

If one asks the first question then perhaps an appropriate response would cite a more ultimate desire to feel good. Here, one would not be asking about the actual cause of my feeling, but rather asking me to tell them why my feelings matter to me. This way of probing might aptly call for the response to cite a desire, something with propositional content. If instead the probing takes the form of Question 2, “why

---

5 To Question 1, “why does it matter to you that it feel good to take care of your children?” it is not clear that I would want to cite a desire here either. I could just as easily respond that the reason it
"does it feel good to take care of your child?", then it seems that what is being asked about is the cause of my feelings—a more specific reason for why taking care of my child feels good, and this more specific reason could be an emotion. As was discussed in the above Appendix, many researchers maintain a distinction between emotion and feeling. For instance, feelings of hunger would not qualify as an emotion. Indeed, Prinz argues that emotions are a particular kind of feelings, whereas feelings represent a more general category of affective states including hunger, fatigue, emotions, tickles and sensations. When inquiring about the nature or cause of a feeling of pleasure in taking care of my child, it seems that what is being requested is a more specific type of feeling, and an emotion is the best candidate here. Feeling hungry or tired or feeling an itch or a headache would hardly seem to constitute reasons for why taking care of one’s child feels good. An adequate response to the kind of inquiry reflected in Question 2 might be something like “I feel love, deep affection, and sympathy for him and so it feels good to help him”—a response that does not cite a prior desire.

Of course even if one could successfully determine that I also, rather plausibly, simultaneously have the desire to feel good, this does not mean that my desire to feel good is the cause of my desire to help my child. Admittedly, there is also another way that a desire could be implicated in causing a feeling, and this is when a feeling results from the satisfaction of the desire. Returning to the construal of Butler’s example, one feels pleasure when a desire for food is satisfied. Likewise, I

matters to me is because I love them. The way of posing the question assumes that there is a desire to feel pleasure that is more ultimate, but it need not be.
could feel pleasure when my desire to take care of my child is satisfied. But of course merely having the desire to take care of my child alone is not enough to account for feelings of pleasure—my desire must also be satisfied. And when we are inquiring into why it feels good to take care of my child, what we want to know is whether there are any reasons prior to having the desire satisfied for why taking care of my child feels good. It would not seem that my desire, by itself, is the reason for my feelings of pleasure. More generally, it seems implausible that desires by themselves cause feelings. To the question “why does it feel good to take care of your children?” it would seem to greatly strain the meaning of the concept of desire to respond “because I have a desire to feel good”. Whether this assertion is true or not, the important point is that it is not necessary that we explain a feeling in terms of a prior (egoistic) desire.

By focusing on alternative meanings behind a citation of internal rewards in my first point I have somewhat shifted the debate from addressing the problem of hedonistic desires. I have only shown that there is a way to cite feelings and avoid the hedonistic trap. Again, I have done this because it seems important to the debate on altruism to show not just that altruism exists, but that it is widespread. In our explanations of altruistic behavior, we frequently refer to our feelings as the reasons, and my intent was to show that this sort of referral alone is insufficient to indicate egoism. The main problem with the Butlerian argument, however, was that if we cite a more ultimate desire to avoid pain as our reason for helping someone, then prior suitability fails to refute egoism. I will address this now by turning to my second, and more important, reason for diverging from the standard view. My response here is
closely related to my first point: the argument against prior suitability, and the view of altruism that underlies it, draws a sharp boundary between feelings and altruism, but, as we will see, feelings are essential to genuine altruism.
2.3.2 Reason II

Once again, the standard view maintains that genuine altruistic motivation consists in an ultimate desire for the improving or protecting the welfare of another. In order to understand what is missing from this view of altruism, it may help to once again look at S&W’s demonstration of the contrast between altruistic ultimate desires and egoistic ultimate desires. As before, the altruist (A) is represented in the first column, and the egoist (E) in the second. But here, I have modified the diagram somewhat in order to capture what was missing in S&W’s construal—namely, the contrast between an ultimate altruistic desire and an instrumental altruistic desire.

Remember that the hard problem of genuine altruism is not in establishing that we have desires to promote the welfare of others, for surely we do. This fact is what makes egoism, and hedonism in particular, so tough to topple from its privileged perch in much of the sciences and social sciences. The problem that proponents of genuine altruism face is to distinguish the desires to help others that a genuine altruist can be said to have from the desires to help others that an egoist can be said to have. In the case of parental concern, A’s desire for her child’s welfare is ultimate. E’s ultimate desire is to avoid pain, but this desire gives rise to her instrumental desire to help her child. So both motivational pathways involve an altruistic desire. We now need to determine whether it is obvious that one of them is more genuine than the other.

In the diagram below, I have represented E’s desire for her child’s welfare as an instrumental one, with her ultimate desire being that of minimizing pain.
If we focus on desires as the ultimate reason why each parent helps her child, then E helps her child because she desires not to feel bad, whereas A helps her child because she desires to promote her child’s welfare. But focusing on the contrast between E and A in this way forces us to miss another important consideration—what triggers the desire to help in each case? Remember from Chapter 1 that the D/I Principle states that given two mechanisms for producing a behavior, the direct one will be more reliable than the indirect one. Because E’s desire to help her child is mediated through her feelings, there is an extra step involved when E helps her child than when A helps her child. So an altruistic desire is a more direct mechanism for helping behavior than is an egoistic desire. This is still the case even with the addition of ultimate desires in the diagram above.

E’s desire to help her child is triggered by a feeling and not directly by a belief. A’s desire to help her child is triggered directly by a belief. But what this
means is that while feelings are a necessary part of E’s motivational path, the same is not true of A’s motivational path. This does not mean of course that feelings are not a part of A’s altruistic motivation. More generally, this does not mean that altruistic ultimate desires do not involve feelings in some way. As S&W argue, certain feelings, namely sympathy and empathy, can sometimes and even frequently be the cause of altruistic ultimate desires. However, they argue that while common-sense suggests that empathy and sympathy trigger altruistic desires, it is still an open question whether the ensuing desire is instrumental or ultimate. Perhaps, they suggest, these feelings trigger altruistic desires only because people do not like having these feelings and want to help others as a means of reducing these feelings. So, according to them, when I learn that my friend is being harmed, I feel bad, and feeling bad is an aversive feeling that I want to eliminate.

S&W recognize the distinction between feeling bad “for someone else”, a sentiment typically associated with sympathy rather than empathy, and simply feeling bad because I learn of some misfortune or suffering. They agree with much of the relevant literature that the latter kind of feeling is more appropriately labeled personal distress. As the name implies, personal distress is focused on the self and involves feeling bad but without feeling bad for someone else. Thus, with the distinction between sympathy and personal distress in place, it might seem as if there is a way to classify a resulting desire as altruistic or egoistic—my desires triggered by personal distress are acquired instrumentally merely so that I can stop feeling bad whereas my desires triggered by sympathy are directly concerned with helping another person. For the standard philosophical definition of altruism, however, there is obviously still
a problem with this response. Even if my feelings of sympathy trigger a desire to help another, there could still be a hedonistic desire lurking in the background that tells me to get rid of sympathetic feelings for others by any means possible. I will return to say more about this below.

Aside from this problem, S&W argue that feelings of sympathy cannot come to the rescue of altruism anyway. This is because they think that these feelings, and others like them, are not a necessary part or cause of altruistic desires. They argue along with other contemporary researchers that feelings and altruistic desires can come apart. For instance, borrowing an example from Nancy Eisenberg, they write “It is possible to enter these emotional states by thinking about problems that have already been solved. Suppose Wendy discovers her husband’s infidelity, divorces him, and then creates a good life for herself. If she then recounts this sequence of events to Walter, he may find himself empathizing with the Wendy of a few years past. What does this empathy motivate Walter to do? Apparently Walter empathizes without forming the desire that Wendy’s situation be improved.” 6 Their point is that feelings for and about another’s plight do not always cause a desire to help. It is easy to see problems with the example S&W have chosen. Even though Walter does not form a desire with the content “help Wendy divorce this guy” or “help Wendy seek counseling to get over him” since these events have already occurred, there are other kinds of desires that Walter could form. He might form a desire about the past, of course, such as a desire that Wendy never had experienced these painful moments, but this is probably closer to the concept of a wish than a desire since it would not be

motivating in the same way. For this example, then, think of a desire that Walter might form about the present, such as to ensure that Wendy really has moved on and to express that he is always ready to listen. These are not trivial altruistic acts. They consume time and energy. Despite being able to conceive of possible altruistic desires that Walter forms in response to feeling sympathy, however, it does seem appropriate to keep the concepts of emotions and desires distinct, as I have discussed before and will say more about at the end of this chapter. It seems we can have emotions of empathy and sympathy without forming an altruistic desire.

But S&W also think the opposite is possible. They argue that we frequently have altruistic desires without any such feelings, such as when we learn of disasters and cruelty in distant places and form ultimate desires that the people in those places fare well.\(^7\) In other words, they think genuine altruism frequently occurs without any feelings whatsoever. This is because they believe that the distinction between instrumental and ultimate desires is all we need. As they rightly note, the distinction between ultimate and instrumental desires “is not a technical and esoteric idea but a concept that is familiar from everyday life. Some things we want for their own sakes; other things we want only because we think they are means to some more ultimate end.”\(^8\) While this distinction seems prima facie convincing as the basis of the difference between egoism and genuine altruism, as we will see, a deeper exploration of the points at issue makes it less plausible that their difference lies solely, mainly, or even at all, in the structure of desires. In so-called detached examples of altruism,

\(^7\) Sober and Wilson, pp. 237
\(^8\) S&W, p. 200. See pp. 200-222 for a discussion of some of the issues surrounding the distinction between ultimate and instrumental desires.
where we form desires about the welfare of people in distant places, it seems true enough that we would not always feel sympathy or any one particular emotion. But some feeling or disposition for a feeling, perhaps of worry or even anger, might accompany or cause the desire to help. So just because sympathy in particular is not experienced, this does not mean that some other feeling is not experienced. In cases where there is no affective component or disposition at all, I am prepared to bite the bullet and say that these desires simply do not represent genuine psychological altruism.

To see why, return to the diagram above. In A’s motivational pathway, feelings do not enter the picture at all, or at the very least, they need not. What triggers A’s desire is a belief. In a similar manner, we could describe many instances of desires that are triggered directly by beliefs alone. When I acquire the belief that it is raining, this might give rise to my desire to wear a raincoat, or from the belief that I need a pen if I want to write I might obtain the desire for a pen, or upon having the belief that I am low on gas I desire to fill my gas tank, and so on. The important point is that nowhere in any of these descriptions is any kind of feeling. Each desire is keyed directly to a belief and not to any other mental state. In the case of staying dry, of writing, and of driving, it might be perfectly accurate to describe the motivations behind these actions as not involving, and never involving, feelings. But is it accurate to describe genuine altruism in this way, as occurring in the absence of feelings, without any role for feelings? I do not think it is. Altruism is different from these other motivations, and not just in terms of representing different kinds of objects. Altruism is a concept in everyday usage that is associated with feelings of sympathy
and love, or with anger and outrage. These feelings for others are what gives altruism its force; without them altruism is merely a desire like any other, a representation of some proposition that we want to be true, such as “water in the glass” if I want there to be water in the glass.

Our everyday sense of altruism involves emotions, and without them altruism is merely a cold representation of something we want to be true. Of course, even if one agrees that certain emotions are a necessary part of genuine altruism, one could maintain that ultimate desires are also a necessary part. According to this line of argument, we would still face the task of determining whether a given altruistic desire is instrumental or ultimate in order to know whether someone is an egoist or an altruist. Once we establish that emotions are necessary for altruism, then, we will want to know just how important their role is: are they the determining factor between egoism and altruism or is the structure of desire, as the objection would maintain, still the most important thing? To answer this question, we need to construct each motivational pathway as a representation not only of the structure of desire but also of the specific, ultimate emotional cause of the desire to take care of one’s child. Specifically, we need to distinguish the role of an instrumental vs. an ultimate desire simultaneously while distinguishing two different emotional causes.

Here we have S&W’s diagram of the two different motivational paths, modified somewhat by leaving out the role of beliefs since they would merely complicate the diagram without making any differences. Since we are trying to determine which parent is the altruist here and it is not a given, I have labeled the two parents X and Y instead of A and E.
Both X and Y are presented as having different emotional causes behind their respective desires to take care of their children. In this scenario, Y feels love for her child, which is of course the common emotion parents experience towards their children. X, however, is amused by her child—her child is an interesting pastime. I have purposively chosen two very different kinds of emotions in order to isolate the role of each emotion from the structure of desires. Furthermore, because of the difference in the structure of desire between X and Y, subsequent descriptions of the role emotion plays in each will be different: because she is amused by her child, X desires her child’s welfare; because she loves her child, Y feels bad when her child is harmed. Even though emotions play a role in both parents’ desires, there is an important difference between their desires for their children’s welfare. Y’s ultimate desire is to avoid pain/seek pleasure, whereas X’s desire for her child’s welfare is ultimate. The way that her emotion gives rise to her altruistic desire is similar to the way S&W discussed the possibility of sympathetic emotions giving rise to other altruistic desires—both then and now, the altruistic desire is ultimate.
Thus, in the diagram given here, the standard difference between an egoist and an altruist is represented without blurring the line between the two. Now, which of the two motivational pathways do we want to call more genuine than the other? Y’s desire for her child’s welfare reduces to another desire, to avoid pain. This much is still true. But Y’s desire for her child’s welfare also reduces, in an important sense, to her love for her child. The cause of Y feeling bad when her child is harmed, and hence the more ultimate cause of her desire for her child’s welfare is the love she feels for her child. On the other hand, X’s desire for her child’s welfare does not reduce to any other desire. However, it does, in an important sense, reduce to something else—her amusement with her child. Between the two, I am more inclined to say that Y is the more genuine altruist. The difference between X and Y can be brought out even more if we treat Y’s desire to avoid pain as falling outside of the causal path. This could be because Y’s desire to avoid pain is a background desire that is not directly activated when she sees that her child needs help, or it could even be, as S&W, think is perfectly possible, an unconscious desire. They write that “the fact that individuals sincerely want to help others, and do not consciously experience this desire as involving a sacrifice or a conflict with their authentic selves, does not tell us what their ultimate motives really are. There is nothing in the egoism hypothesis that prohibits other-directed desires from being fully integrated into the agent’s personality. Those desires must be instrumental, but people need not experience them as alien intrusions.”9 If Y’s desire to avoid pain is not experienced but her love for her child is, and if what is experienced for X is her amusement with

9 S&W, p. 226
her child, then there seems to be an even larger difference in the genuineness between X and Y’s altruistic motivation. Even if Y’s desire to avoid pain is not unconscious, though, the point remains that in terms of the emotions felt, the more ultimate motivation for her helping her child is her feeling of love, whereas the more ultimate reason for X to help her child is her feeling of amusement.

An immediate objection to what I have just demonstrated might be that if X’s desire for her child’s welfare is caused by her amusement with her child, then this means that this desire would not be an ultimate one. Instead, a desire for amusement would be ultimate—her feelings of amusement could be viewed as being constitutive of a desire for amusement or a desire for amusement could get activated by feelings of amusement. Either way, her desire for amusement would be more ultimate than, and would cause, a desire for her child’s welfare. Her desire for her child’s welfare then would be an instrumental one. This might be a plausible description under some conceptions of emotions, such as those sometimes attributed to Hume, in which feelings and desires are brought closer together. However, in many accounts of emotions, such as the somatic feeling and behaviorist views discussed above, feelings and desires are distinct. If feelings and desires are thought to be distinct, it is possible that X’s amusement with her child would give rise to an ultimate desire for her child’s welfare and not first, or at all, to a desire for amusement. As I discussed in the previous chapter, it does not always seem to be the case that a given feeling gives rise to a specific desire. The feeling of being angry at someone may give rise to a desire to seek revenge or it may give rise to a desire to kick a shoe or even avoid that person. Thus it is plausible to maintain that feelings of amusement would give rise to
X’s desire to take care of her child, without first forming some other desire. Of course one might insist that feelings of amusement do not give rise to ultimate altruistic desires, whereas feelings of sympathy and love always do. I hope to demonstrate, however, that even granted this conceptual pairing between certain feelings and a particular level of desire, it would still be the feelings that would be doing the work in our understanding of genuine altruistic motivation and not the level of the desire. Furthermore, it does seem plausible, as S&W have argued, that feelings of sympathy, at least, would sometimes bring about only instrumental desires. Thus it is not clear that this conceptual pairing should be granted.

If one adopts the perspective that desires and feelings are not distinct, but rather two different components of the same mental state—one conative, or motivating, and one affective—there are even better reasons for maintaining that feelings play a more prominent role in our understanding of genuine altruism than does the level of the desire. As discussed before, feelings for others are what gives altruism its force; without them altruism is merely a desire like any other, a representation of some proposition that we want to be true, such as “water in the glass” if you want there to be water in the glass. Notice that even a regular old desire like wanting water raises the question “why do you want water?” In other words, why do you want the proposition “water in the glass” to be true? It is probably because you are thirsty or care about your health. But even if I do not want to know more about your desire for water, I would not doubt the authenticity of your wanting water. The genuineness of your desire is not something that would be at issue. Of course, the
same is not true of altruism, since the central debate about altruism is about
distinguishing genuine altruism from apparent altruism.

The search for genuine altruism is an inquiry about fundamental motivations.
The point of trying to discover whether a particular altruistic desire is ultimate or
instrumental is to figure out what, at the core or sincerely, motivates someone to help
someone else. We want to know what they really care about and we will not be
satisfied until we have gotten to the root of the motivation. But why does this
motivation have to come in the form of desire? Furthermore why does inquiring into
what someone really cares about imply that only an ultimate motivation will suffice
in the response? In terms of the first question of why an altruistic motivation must
come in the form of a desire, we saw before that feelings can motivate. This is not to
suggest that feelings are always motivating, or that all feelings are motivating. All we
need is the more modest thesis that certain feelings are motivating at certain times, or,
even more minimally, that feelings give rise to desires and these desires are what
motivates. To be clear, I am not arguing that the idea of altruism would make more
sense if we only included affect apart from any conative factor, whether a desire or
something else. Even if one were to insist that feelings do not motivate on their own
and thus that they require, in addition, an altruistic desire, my concern is whether the
fact that a desire is ultimate is what determines whether an altruistic motivation is a
case of genuine altruism or whether, alternatively, it is the affective component that
determines this.

When we inquire into the cause of some altruistic action, there are at least two
different ways of posing the question. Thus far, the discussion of altruism has begun
with the question of ‘what does one ultimately desire or want?’ But there is another way of posing the question of genuine altruism and it is to ask, again, ‘what does one really care about?’ To see that they are different kinds of queries, consider an example that is irrelevant to the altruism and egoism context. If you tell me that you have an ultimate desire for doing some trivial activity, such as picking up a pen, not for the purpose of writing anything, but only for its own sake, I would not think that this is sufficient for indicating that you really care about the pen or about picking up the pen. I would know that you desired to pick up the pen for its own sake, but whether this meant that you really cared about the pen or not would seem to be an open question. It seems the latter is our concern in the context of genuine altruism and the pen example illustrates that it would not be satisfactorily resolved by citing an ultimate desire. I will return to say more about the insufficiency of citing an ultimate desire in the next section.

If the question of genuine altruism is ‘what does one really care about’, then a response that cites a feeling would seem to satisfy the query. In the first place, caring is often something that the subject does consciously, not something that is unconscious or hidden to the subject. Frequently, we would not say that you care about some object if you had no awareness of your feelings of concern for that object, even if you take actions on behalf of it. Say that you work very hard and as a consequence achieve great success in your career according to some material standards. If you have no interest in being ambitious in this sense, and instead sincerely deny ever being concerned about your material success, it would be odd to insist that you really do care about your material success but you simply are not
aware of it. In fact, we say that people who do not ever consciously consider their material success in their careers do not really care about being successful in that way. We would say that other things motivate them, perhaps love for the work or perhaps that they do not love the work but are just really good at what they do.

Of course in many cases it seems likely that people can come to discover that they care about something by finding themselves acting to protect it, even if they were previously not aware of any feelings of concern for it. However, despite occasions where people lack awareness of feelings and yet still seem to care about the object they are acting for, there is a more relevant point for the purposes here. Cases where people do experience feelings of concern would also seem to be cases of caring. In fact, the experience of feelings of concern for an object while taking actions on behalf of it would seem to be sufficient for caring. There are a variety of examples that would support this, but to take just one, consider someone working for Greenpeace who feels great passion for the environmental work she does. In this case and others like it, it seems obvious the person ‘really cares’ for the object in question.

It seems that really caring about something frequently requires, or at least would be satisfied by, a conscious mental state—something with a phenomenological feel—even if only from time to time, such as when one ponders the object in question or reaffirms an interest in it after moments of doubt. But if this is the case, then the best candidate for indicating whether someone cares about something is a feeling, because a feeling just is something a subject is aware of. We saw earlier that S&W and other philosophers maintain that we can have unconscious ultimate desires.

---

10 Allen Buchanan made this very helpful point to me.
lurking in the background, but it is far less plausible that we could have unconscious feelings. To take a very general example of a feeling, consider pain. We would not insist that a person is in pain if she does not feel pain. It is the experience of pain which defines pain. We would say that this person has suffered an injury but does not feel it or at least does not yet feel it. For an example of a feeling that is tied with emotions, consider someone who loses her child. If she is in shock and sincerely reports not feeling any sadness, then it would not make sense to insist that she does in fact feel sadness but simply does not know it. It would be more appropriate to say that she does not feel sad yet because she has not internalized or accepted what has happened—that she is still in shock. While we can be mistaken about or unaware of our ultimate desires even after sincere reflection and under normal circumstances, this seems far less true when we consider our feelings precisely because they do have a phenomenological feel. This is what they consist in. Thus, while a sincere avowal of a feeling seems to be sufficient for understanding what one really cares about, a sincere avowal of an ultimate desire does not seem to be.

2.3.3 Reason III

If our interest is in getting to the psychological heart of the matter, and we want to know what someone really cares about, then it does not seem as if we will be satisfied with knowing that one has an ultimate desire for another’s welfare. A critic of what I have just argued might respond that the question of ‘what does one really care about’ can be answered satisfactorily with the citation of an ultimate desire
because of what a desire being ultimate is meant to indicate about genuine motivation. What we must do then is examine precisely what the appeal to ultimate desires on the part of the theorist is supposed to accomplish and whether it succeeds in this task. It seems there are three different criteria of genuine motivation that ultimate desires are thought to satisfy. These are stability, strength, and uniqueness of object. However, I will try to demonstrate that ultimate desires do not necessarily satisfy these criteria for genuineness, giving us our third reason for rejecting the standard view of altruism.

The main philosophical account of ultimate desires maintains that they are more stable than instrumental ones. S&W elaborate this position most clearly by arguing that the important difference between an instrumental desire and an ultimate one is that when I desire something only as a means to some other end, I am more likely to give up the means in light of new information. When I learn that my instrumental desire no longer brings about the satisfaction of my ultimate one, I will give it up. In addition, anything that causes me to give up my ultimate desire also causes me to purge my instrumental desire, while the converse is not true: when I lose an instrumental desire this says nothing about the ultimate desire that it served. Theorists do not think that ultimate desires are impervious to change, only that instrumental desires are much less impervious to change. Likewise, we can add that it seems that an ultimate desire would be stronger for related reasons. One way in which it would be stronger than the corresponding instrumental desire is if I were to have reason to doubt the success of the instrumental desire in causally contributing to the ultimate one. For even if I do not purge the instrumental desire from my set of
desires but only lose my faith in it, I may cease at any moment to be motivated by it, or, at the very least, its motivational strength may decrease, while I would still be strongly motivated by the ultimate one. Thus, ultimate desires can be said to be both stronger and more stable than instrumental desires. But what this view assumes is that the linguistic structure of desire is equivalent to psychological strength and stability and the latter does not always seem true.

I could easily have some very stable instrumental desires. Consider having the desire to eat food only because it promotes my ultimate desire to live or only because it promotes my ultimate desire for pleasure. Even though it would be an instrumental desire, because of the causal importance of food to the end of living or experiencing pleasure, and since the latter two are significant ultimate desires, my instrumental desire for food would be very stable. Conversely, I could also easily have very fleeting ultimate desires. Imagine the desire to read political cartoons, not because of something else it promotes but as an end in itself. I simply like reading political cartoons. But if I learned of a new pastime that I liked better I could lose this ultimate desire altogether. Now compare my ultimate desire to read political cartoons to my instrumental desire to publish in philosophy. I do not try to publish journal articles in philosophy for the sake of publishing itself, but because it promotes my ultimate desire to engage in the discourse of philosophy and to be a professional philosopher. Even still, it is a much more stable desire for me than my ultimate desire

---

11 Even though we might be able to say that one gets pleasure out of reading the cartoons, we should resist the temptation to recast this desire to be an instrumental one. If we do that, then we could do the same to ultimate altruistic desires and then there really is no philosophical debate worthy of our attention. In order to take the concerns in the literature seriously, we need to be able to treat certain desires as ultimate without the temptation to reduce them to other desires and so this is also true of the desire to read political cartoons.
for reading cartoons because what publishing philosophy is instrumental to, namely, my desire to be a philosopher, is a very stable desire, and even more importantly publishing philosophy is causally very important to being a professional philosopher.

This last comparison brings out an important point. How we view the stability and the strength of various instrumental and ultimate desires depends on the mode of comparison. Compared to the ultimate desire that an instrumental desire serves, the instrumental desire will probably always be weaker: my desire to publish philosophy is weaker than my desire to be a professional philosopher. Even this fact might not obtain if an instrumental desire is linked to more than one ultimate desire, as with the example of eating food which serves both an ultimate desire to live and one to experience pleasure. Because it serves two ultimate desires, I could lose one of them and still retain my instrumental desire for food, rendering it more stable than each of the ultimate desires it serves. Even more telling, if we take an instrumental desire, such as my desire to publish philosophy, and compare it not to the ultimate desire it serves but to another ultimate desire, there is no reason to think that the ultimate desire will be stronger or more stable: my instrumental desire to publish philosophy is stronger than my ultimate desire to read political cartoons. I may sometimes experience weakness of will when it comes time to taking actions on behalf of publishing philosophy and would momentarily prefer to read cartoons, but my desire for the latter would not be stronger than my desire to publish philosophy all things considered.

Likewise when it comes to altruistic desires, it is easy to imagine that an instrumental one is stronger than an ultimate one. Picture a creature slightly less
cognitively developed than humans. Since he has not faced all of the environmental complexities and variations that humans have, he has never developed more than a single ultimate end—to avoid pain and seek pleasure. This hedonic creature still cares very much about the welfare of his child because this is related to his ultimate hedonistic desire. In fact, given the limited environmental and social variability, there are only a few other instrumental desires that this creature has, such as desires about food, sex, physical pain, hierarchy in his social group, and attachments with other conspecifics. Now consider that S&W and other philosophers are right in their conception of genuine altruism. As a human being, all of my genuine altruistic desires are ultimate ones, but because of the complexities of our environment and culture, I have many such desires. Among my many ultimate altruistic desires, I have a desire to help strangers find their way in my city. It seems obvious to consider that the hedonic creature’s instrumental desire for the welfare of his child is much stronger than my ultimate desire to help strangers with directions. If you are unhappy with the comparison to another kind of creature, imagine that your only ultimate desire is to avoid pain and seek pleasure. One of your main instrumental desires is to take care of your friends—their harm brings you great pain. Compared to your ultimate hedonistic desire, your desire for the welfare of your friends will be weaker. But when compared with my ultimate desire to help strangers with directions, it is much less plausible to think that your desire for the welfare of your friends is weaker. To represent this point more generally:
The double arrow indicates both that (A) gives rise to (a) and that (a) satisfies (A). (a) would be weaker than and less stable than (A), but we cannot infer any relationship about strength or stability between (B) and (a). 12 For all we know, (a) is stronger than (B). If we had more information about these desires we might be able to say that (B) is stronger than (a). But the point is that there is nothing in the linguistic representation of these desires itself that tells us this. It might be the case that in general, or on average, ultimate desires are stronger and more stable than instrumental ones, although even this is far from obvious for the reasons given above. Even if true, however, the tendency for ultimate desires to be stronger or more stable does not tell us anything about a particular ultimate desire. If I were to show you that people tend to spend more as their income increases, this would not allow me to say anything about the spending habits of a particular individual. While we might want to say that spending frequently depends on income, it would not allow us to define

---

12 As discussed above, even the first point can be questioned. If (a) is only instrumental to (A) without being related to any other ultimate desires, then it seems (a) would certainly be weaker than and less stable than (A). However, if (a) is an instrumental desire that serves additional ultimate desires, then it could be stronger and more stable than (A).
spending in terms of an income increase or even a qualified kind of spending, such as ‘serious spending’, as long as there are cases where the two factors, spending and income, come apart. Insofar as the claim that ultimate desires are more genuine than instrumental ones depends on claims about stability and strength, this is an empirical claim. Since it is empirical and, as it turns out, not always true, it does not make sense to make the claim the defining aspect of genuineness.

I argued that the structure of desire in itself tells us little about whether a desire is stronger or more stable. But there is another way of thinking about genuineness and this is to say that the object of an ultimate desire is special and unique. The object of an ultimate desire is irreplaceable with other objects; for the only thing that satisfies the desire is this object. On the other hand, because instrumental desires are substitutable, so are the objects of these desires. This idea is intuitive but the notion of irreplaceability is not limited to the objects of desires. It is not the case that the only way to gain this criterion of genuineness is through distinctions in the level of desires. We can gain it in two other ways. The first way is generalizeable to all kinds of desires, including but not specific to altruistic ones, because it involves the relationship between desires. Consider again my desire to publish philosophy that serves my ultimate desire to be a professional philosopher. Because the desire for publishing philosophy is instrumental, if I discover that writing fiction would better serve my end, I would discard the desire to publish philosophy. But, and this is the point, I doubt I would discover this. There is something about publishing philosophy that makes it qualified as a way of bringing about my end of engaging in the discourse of philosophy. Through rejections, reviews and acceptance,
trying to publish hones one’s skills at presenting philosophical ideas and reasoning through arguments. And public accessibility to one’s philosophical work invites discourse with others over time which in turn allows one to reflect further on the ideas. In this way, publishing philosophy is unique. It is uniquely related to my end of being a philosopher.\textsuperscript{13} Of course one could argue that my desire for reading political cartoons is even more unique because of its status as being ultimate. While it is true that the object of an ultimate desire is special and unique in the sense that it is the only thing that satisfies the desire, this might shift the burden of establishing uniqueness to the desire. How strong is this desire? Over what period of time is it unique? How many such ultimate desires does one have? If one has countless ultimate desires, would we still consider the object of any one of them to be unique or special?

My desire to read political cartoons is a desire that I can easily imagine discarding, however unique, if I discover some other pastime. In a way, its uniqueness contributes to the possibility that I will lose interest in it. I might get sick of it. While my desire for reading political cartoons can be characterized as very unique for a limited time, my desire for publishing philosophy is somewhat less unique for a much longer period of time (my entire career) and is stronger than the former desire.\textsuperscript{14} But perhaps it would make sense to assume that people have very

\textsuperscript{13} I do not want to say that publishing philosophy is constitutive of being a philosopher, because I want to retain the distinction between instrumental and ultimate desires. But if the reader is not convinced with this distinction, she can substitute this example with my earlier one in which eating brings about pleasure.

\textsuperscript{14} To put the comparison in terms of romantic relationships, my desire for reading political cartoons is more like a short-lived infatuation, whereas my desire for publishing philosophy is more like a committed, happy marriage. I am happy to say that the latter feels more genuine.
few ultimate desires—not for things like pens but for objects such as one’s well-being and one’s family. Thus the object of an ultimate desire is thought to be unique because ultimate desires are rare. However, why should we assume that ultimate desires are rare? The most likely reason would be that there is a psychological limit to the number of objects one can consciously care about, and that an ultimate desire is said to represent something one consciously cares about. This is not of course the common way in which philosophers describe ultimate desires. More importantly, even if they were described in this way, as discussed under Reason II, this would seem to bring us back to making feelings the defining feature of caring, and not the level of the desire.

My comments about gaining the quality of uniqueness through the relationship that exists between a particular instrumental and ultimate desire apply to motivations of all kinds. There is another way that we can gain this criterion of genuineness which applies only to altruistic motivation. Our everyday intuitions on altruism might suggest that an ultimate altruistic desire is required because it is a desire about or for another person that does not reduce to any desire that is about or for oneself. Since I have already discussed the insufficiency of desires alone constituting our notion of genuine altruism, the important consideration for this way of describing genuineness is the aspect of another person. Unlike the uniqueness quality that can be said to exist in any context in which an instrumental desire uniquely satisfies a particular ultimate desire, now I want to focus on uniqueness of a particular sort—the distinction between oneself and another person. To represent your uniqueness from me, my motivation to help you must be about you and not me.
While it is beyond the scope of this project to discuss the different range of feelings that are associated with altruistic motivation, I wish to point out that we can achieve this element of a motivation being about someone else and not the self through differences in feelings. To take a prominent feeling identified with altruism, as I discussed above there is an important difference between sympathy, which can be described as feeling bad for someone else, and personal distress. Personal distress as with sympathy involves feeling bad when I learn of or witness some misfortune or suffering. But, as the name implies, personal distress is focused on the self and involves feeling bad without feeling bad for someone else. On the other hand, sympathy is about another person and not oneself. It is no coincidence that sympathy has been closely associated with altruism both historically, through the work of Hume, Smith, and Butler, and in our everyday lives.

As I described in my first reason for disagreeing with the standard view of altruism, if one cites a feeling like sympathy instead of a desire as her motivation for helping another, it does not make sense to inquire into the desire that caused the emotion. There might be other kinds of causes we would want to know about of course: did feelings for the other emerge from close contact over the years; are they a result of being members of the same group or community; are they due to shared similarities or simply because the other individual is a being capable of suffering, etc. However, as argued above, it would greatly strain the concepts of ‘desire’ and ‘feeling’ to inquire into the desire that caused a feeling. But there is an asymmetry here in terms of explaining altruistic motivation. We will frequently want to know the emotional causes of altruistic desires. Altruistic actions are not trivial ones like
picking up a pen. They are by definition costly to the self. For this reason, we do not just want to know what someone’s desires are, but, why they have the desires they do. The altruistic parent from before, A, had a desire to help her child that was triggered by a belief. But, without more, her altruism was merely a cold representation of something she wanted to be true. We will want to know why A has a desire to help her child—in other words, whether she really cares about her child. It seems we will only be satisfied in our inquiry if A feels love and affection for her child, or more generally, that A feels sympathy and compassion for suffering others. Otherwise it would not represent what we common-sensically see as genuine altruism.

2.4 Conclusion

To conclude this chapter and the previous one, I have discussed how the standard conception of genuine altruism is misguided because it excludes any essential reference to feelings. I have also argued that the search for genuine altruism is an inquiry into what someone genuinely cares about, and that we have options other than determining whether a desire is instrumental or ultimate for understanding the notion of genuineness. Ultimate desires for the improvement of another person’s welfare were said to be more genuine because of what they are thought to indicate about motivation—strength, stability, and a unique object. However, I have argued that whether a desire is ultimate does not always tell us much about these criteria. What this means is that an ultimate desire fails to satisfy those criteria which were said to render it a satisfactory response to the question “what does one really care
about?” On the other hand, I have attempted to demonstrate that citing certain feelings such as sympathy would satisfy this question because caring is something that implies the experience of feelings.

The contemporary effort to distinguish altruism from the pursuit of internal rewards has divorced genuine altruistic motivation from feelings, but feelings are important to both: (A) what it means to genuinely care about something, in general; and especially, (B) the notion of altruism, as understood both historically and intuitively. Furthermore, ultimate desires do not seem to be necessary for genuine altruism. To put all this together, my conclusion is that certain feelings are sufficient to indicate a case of genuine altruism. As discussed in my introduction to the previous chapter, my intent here has not been to establish the particular range of feelings that are important for altruism. I have rather tried to demonstrate that something important is missing from the standard view, and this is the role of feelings. Of course, I have not specified the extent that feelings operate in a notion of genuine altruistic motivation. Instead, the goal was to clear away the conceptual baggage that has notoriously prevented both an acceptance in rational choice theory that genuine altruism exists and efforts to theoretically and empirically explore the reach and context of altruistic motivation—something I begin to do in the Appendix following this chapter. Once we account for the prominent role of feelings in understanding altruism, and at the same time recognize the limited, and perhaps absent, role of ultimate desires, it can free us from the hedonistic trap mentioned before.
2.5 Appendix 2: Applications to Rational Choice

In this final part of my discussion of altruism, I extend my analysis of the standard view of altruism to discuss some specific applications to rational choice theory. For the sake of continuity with the terminology of ‘preference’ rather than ‘desire’ used in the rational choice literature, in what remains I will refer to the standard view as the ‘preference view’ of altruism. I will first explore the implications of my proposed way of understanding genuine altruism to some theoretical issues and then discuss how this view of genuine altruism might bare on empirical efforts to test for the prevalence and conditions of altruism.

2.5.1 New Avenues for Theoretical Research

Most of rational choice theory includes the notion of ‘revealed preference’ which maintains that choices reveal what someone’s preferences are. A common criticism of revealed preference theory is that the theory does not allow that individuals’ choices and behavior will sometimes, perhaps frequently, be irrational or fail to reflect their ‘true’ or reflective preferences.15 A further problem with revealed preference theory is that it does not allow us to explain behavior in terms of preferences—to say that preferences cause behavior. But even if we keep the concepts of preference and choice distinct, there is a further concern in the context of altruism; as explained before, the preference view of motivation makes no essential

---

15 For a brief discussion of some of the concerns with RPT, see Hausman, Daniel and McPherson, Michael. 1996. Economic Analysis and Moral Philosophy. Cambridge: Cambridge University Press, pp.27-29
reference to the role of emotions and therefore underplays the significance of feelings in altruism.\footnote{I explore this issue much more extensively in my Dissertation}

Even though I have been arguing that the question of whether a desire or preference is ultimate is misguided, and perhaps irrelevant, there is another kind of reduction that should be avoided, and this is the collapsing of emotions to preferences. When rational choice theorists and especially economists have discussed emotions, including altruistic ones, they have typically viewed them in terms of costs and benefits to be weighed against \textit{material} costs and benefits. Emotions are pains and pleasures to be minimized or maximized respectively. They have simply been added to the utility functions as objects of maximization and subject to marginal utility analysis. While the modeling of emotions in this way may be useful for some purposes, it seriously masks the seemingly important differences between the emotions and material interests. As Elster has powerfully argued, the passions, which involve emotions, and material interests are not always fungible.\footnote{He provides an example of an agent deliberating between stealing an expensive book from the library. On the one hand, the individual has material interests in stealing, and on the other, he wishes to avoid guilt because it is an adverse feeling. If guilt were nothing but a cost than the agent should be willing to buy something like a guilt-erasing pill if it was cheaper than the cost of the book. But Elster writes “I submit that no person who is capable of being deterred by guilt would buy the pill. In fact, he would feel guilty about buying it.” (p. 303). A person willing to take the guilt-erasing pill would not need it. (1999. \textit{Alchemies of the Mind: Rationality and the Emotions}. Cambridge: Cambridge University Press)}

Albert O. Hirschman explored the development of the idea that the pursuit of interests can moderate destructive passions, such as glory.\footnote{Hirschman, A.O. 1977. The Passions and the Interests: The Political Arguments for Capitalism before its Triumph. University Park, PA: Pennsylvania State University Press.} A society in which people are motivated by material interests as opposed to passions would be more predictable and stable. Cooperation would be facilitated because of common
interests; people would have the incentive to work together to protect their interests. If the passions could be reduced to material interests, then passions such as glory could be assuaged and pacified by the acquisition of material goods. Hirschman’s analysis implicitly assumes a sharper distinction between the passions and material interests.

There are also other differences between material interests and emotions that are veiled in attempts to model the emotions. As discussed in the Appendix to Chapter 1, a range of emotion researchers, from William James and Karl Lange to Antonio Damasio and Jesse Prinz, have identified the emotions as involving patterns of bodily changes that, through their negative, positive, or mixed feelings, urge us to act. Emotions have a phenomenological feel. Furthermore, I argued before that they are often presented to us as a force or urge so strong that they overwhelm every other concern or interest. We could not easily capture the latter difference by modeling the emotion as lexically prior to other interests and concerns, however, for at least two reasons. First, the overwhelmingness of the emotion may be very temporary, and second, the emotion under such circumstances is not just a cost or benefit in its own right but becomes a lens from which we evaluate the other costs and benefits, the other interests and concerns. It colors, whether for the better or worse, our assessment of what our interests are and what weight we assign to these

---

interests. Hence, passions such as glory would seem to resist substitution with material interests.²⁰

If emotions that might appropriately be characterized as egoistic should be viewed as separate categories from material interests, and thus not capable of being captured in terms of preferences, then it is even more plausible to maintain this distinction when considering altruism. When we focus the discussion of emotions to just altruistic emotions, there is an additional and significant distinction between emotions and preferences that is collapsed in formal models—this is the distinction between the self and another. At least in the case of the status passions, for instance, such as glory, vanity, and ambition, the emotions involved are about the self in some important ways. But, as discussed before, emotions involved with altruism are about others. Representing altruism in utility functions would collapse this further important distinction between something that is about oneself and something that is about another. The utility that someone would derive from helping others would be substitutable with the utility gained from promoting one’s own self-interests. We would not know why one gains utility from the promotion of someone else’s welfare or consumption, and therefore we would not yet understand why there is a preference to help another—is it because one cares about the other or is it because altruistic motivation is a means of attaining political or social acclaim, or some other reasons?

²⁰Rousseau’s view seems to have been that the satisfaction of passions, even those that might be characterized as egoistic, is not interchangeable with the satisfaction of interests. The feelings associated with *amour-propre*, or relative status, for instance, are aimed at bringing harm to others, not at bringing benefit to the self, which implies that they are not reducible to interests. (Rousseau, J.J. 1964. "On the Origin and Foundations of Inequality Among Men". In R.D. Masters (Ed.), *First and Second Discourses*. New York: St. Martin's Press) For an excellent discussion of Rousseau on this point as well as Hobbes, see Grant, Ruth, "Passions and Interests Revisited: the Psychological Foundations of Economics and Politics", unpublished manuscript.
For moral and political behavior at the very least, we should want a richer description of our psychology, and there will be at least some circumstances when these differences matter. Economists’ efforts to model altruism as externalities in utility functions or as mere arguments in a cost-benefit calculus do not go far enough in terms of distinguishing the self from another and are not capable of adequately representing altruism as involving feelings.\(^{21}\)

As discussed above, emotion researchers describe feelings as being central to emotions. Feelings, through their either negative or positive valence, propel or compel us to act. For instance, the feeling of fear would represent a dangerous situation through registering changes in the body, but would also represent the danger as something that is bad or negative, and serve as an inner reinforcer to withdraw from the dangerous object, to eliminate the situation, often in a spontaneous fashion. Because they urge us to want more or less of an object or situation, emotions sound a

\(^{21}\) Robert Frank has made the emotions central to an account of altruism, but his identifying the emotions with altruism is problematic. There are two ways to understand the role emotions are playing—either in terms of their consequences or in terms of being motivations. If we view them in the first way, then all of the emotions in his view, since he takes an individual selectionist framework, contribute to the long-term or ultimate material interest of the person—from this perspective all emotions are egoistic. If we view them in the second way, all emotions are motivations that compete with motivations to pursue short-term self-interest. But in viewing them this way, as competing with self-interest, emotions are all altruistic (or moral, which he takes to be synonymous with altruism). Frank could argue that there can be egoistic feelings, arising from rational calculation. Indeed he seems to hold this belief. In this case, emotions could be either altruistic or egoistic. However, in order for Frank’s account of emotions as commitment-solving devices to succeed, either emotions cannot be egoistic or else egoistic emotions must be sufficiently weak as to not really count in the same category as the altruistic emotions. This is because in the short-run, taking egoistic actions also has short-term material gains on its side—such as those that come from cheating a business partner. In order to counteract both (1) egoistic feelings and (2) material gains that come from taking a short-term egoistic action in the reward system that Frank identifies, altruistic feelings would have to be systematically very strong, so strong that it might not make sense in holding egoistic feelings and altruistic ones in the same category. If altruistic emotions were not systematically much stronger than egoistic ones, then these altruistic emotions could not hope to solve the commitment problems. (Frank, Robert H. 1998. Passions Within Reason: The Strategic Role of the Emotions. New York: W. W. Norton and Company, Inc.)
lot like preferences. Even though above I concluded, in agreement with many researchers, that emotions are, among other things, action-tendencies, there seems to be an important qualification to add. Emotions seem to incite or compel us to act, and in this way can be classified as action-tendencies—as motivations; but emotions do not seem to have aims or intentions. Emotions are *about* something and so it makes sense to describe the objects of emotions, and they do compel and motivate us towards action. But they do not necessarily have an *aim* and especially do not always have the aim of being realized or satisfied in the way that preferences do. Another, but slightly different, way to put this is in terms of the notion of ‘direction of fit’. Preferences have a direction of fit that emotions do not possess; the former aim to bring the world into fit with them.²² For instance, a preference to help someone has the aim to bring about a state of affairs in which that person is helped. The emotion of sympathy, on the other hand, does not aim at any state of the world, whether it is to help someone or even to feel sympathetic.

So far in this section, I have been speaking of passions interchangeably with emotions, but they are not synonymous. While it might be natural to view both the emotions and the passions as picking out the same type of phenomena, the description of the emotions that emotion-researchers provide us suggests that at least many of the political passions are combinations of particular emotions with particular preferences. Plato, for instance, characterized vanity, ambition, and glory as some of the passions that drive political life, but they also aimed at certain consequences. These aimed, for

---

instance, at gaining recognition and victory. In the Rhetoric, Aristotle suggests that emotions involve both affective feelings, such as distress and pleasure, and desires for action. Anger, for example, may be a distressing desire for revenge. Other passions, however, such as shame, seem to have more kinship with the contemporary way emotions are conceived. The terminology is not relevant for my purposes, but again it is important to maintain the distinction between feelings, which are a central aspect of emotions, and preferences. Feelings may frequently give rise to preferences, which in turn have an aim at some behavior or choice, but as I argued above, much of the time one feels angry one might not acquire the preference for revenge. Contrary to preferences, feelings do not seem to have specific goals. I have now added an additional claim: feelings do not seem to have any goals or aims whatsoever. This means that it does not make sense to say that feelings can be satisfied in the same way that preferences can. This claim gains support from the fact that feelings of anger can persist even after someone has achieved revenge or any number of behaviors. Feelings are not exhausted by being ‘satisfied’, but preferences are.

Even though feelings do not necessarily have a particular goal, such as revenge, I argued before that they are often commands to do something, for instance, to curse or kick a shoe. This aspect to feelings is important. Although there is no specific goal or any goal at all associated with a given feeling, they often compel us to act in some way or another. Because feelings do not aim to be realized or satisfied, but are commands to do something, they allow us to make better sense of behavior.

---

that is expressive rather than consequential or instrumental. Expressive behavior is behavior that does not always aim to achieve any outcome, at least at any given point or in any given circumstance. For instance, someone who feels bad for others will sometimes be compelled to just express his feeling bad, even if no one observes or witnesses the expression. He may not aspire to affect outcomes on every occasion. Expressing one’s feelings without being able to change things can cultivate these feelings and foster behavior on other occasions that would be aimed at improving the welfare of others. Thus, the expression of feelings certainly seems to have some relationship with particular consequences, but expressive behavior does not seem to stand in a deterministic relation with particular outcomes, and for this additional reason, would not be subject to cost-benefit considerations.

Especially if we view preferences in terms of choice, as revealed preference theory does, there is a gap between preferences and feelings. We could of course still say that expressive behavior aims to achieve expression, and thus that someone might have preferences to express oneself, but it may not be illuminating to do so. How could we meaningfully determine the costs and benefits of expression when it is not for the purpose of achieving any result? What standards would we employ for measuring costs and benefits? Furthermore, as discussed above, there is sometimes an overwhelmingness to feelings, so much so that they are not just, if at all, objects of maximization, but they also shade our evaluation of other costs and benefits.

Feelings, by causing expressive behavior, thus seem to have many applications to domains of life involving family, community and identity—in other words, those areas where we would expect to see a greater prevalence of altruistic
behavior. Importantly, there is little or no temptation to free ride if what I am compelled to do is express my sense of solidarity, outrage, or sympathy—nobody else can do these things for me. To take just one topic, consider the application of altruistic feelings to voting in large elections. Because the costs of voting to the voter seem to far outweigh the benefits he receives, voting behavior has famously caused problems for rational choice theory. To explain voting behavior, we could of course ascribe altruistic preferences to voters. In these explanations, instead of just seeking their own welfare, voters would have a preference for the welfare of others. But merely assuming an altruistic preference would not be sufficient—the benefits, albeit to others, would not outweigh the costs to the (altruistic) voter. It would still be the case that the marginal impact of any particular voter’s participation, whether on private or public interests, will be very low. A central way researchers have addressed this potential problem is by describing the voter as having a preference to express his sense of civic duty or concern.\textsuperscript{25} While this kind of approach is promising, it fails to offer a satisfying explanation.\textsuperscript{26} These explanations, by ascribing preferences for expressive behavior, maintain the position that voters maximize some kind of expressive value—they are still aiming at a result. Voting would still be subject to cost-benefit calculation, and would suffer from the problem of meaningful standards mentioned before. Furthermore, there is no reason to believe


\textsuperscript{26} For a seminal discussion of how rational choice theorists have frequently accommodated empirical anomalies by expanding the range of preferences to the point that theories become tautological, see Green and Shapiro (1994, \textit{Pathologies of Rational Choice Theory; a Critique of Applications in Political Science}, Yale University Press: New Haven, especially pp 47-53 and 87-97).
that the calculus would come out in favor of expressing oneself, unless we simply
“stipulate” that the benefits of expression will be very high, enough to offset costly
voting. But, again, the important question at that point would be why the benefits of
expression are thought to be so high.

If we do not always view altruism in terms of preferences, however, then the
motivation for voting need not be entirely, if at all, consequential. A voter may
simply be compelled to express his sympathy or concern for others and since he is not
attempting to maximize his expression, it would not be subject to cost-benefit
considerations. Of course, this is only a potential explanation and would still
mandate empirical confirmation if it is to be taken seriously. This last point brings
me to some remarks about the possibility of testing for altruistic motivation.

2.5.2 New Avenues for Empirical Research

I have argued that the effort to distinguish altruism from the pursuit of internal
rewards has divorced altruism from feelings; feelings are an important part of
altruism that may illuminate some significant kinds of behavior. Furthermore, I have
argued that altruistic feelings cannot be collapsed into and captured by preferences.
An important advantage that potentially follows from this last point concerns the
viability of testing for altruistic motivations.

27 Green and Shapiro demonstrate that the Riker and Ordeshook attempt to explain voting by adding a
preference for performing one’s civic duty just stipulates that the psychic rewards from voting are
greater than the costs. Likewise, adding preferences for expressing altruistic concern would seem to be
culpable of similar stipulation.
Although altruism is often researched and discussed across many disciplines, a range of definitions and conceptions are employed and many studies designed to test for altruism fail to make important distinctions. For instance, most research in economics, political science, and anthropology does not distinguish between altruism as a behavior and altruism as a motivation; donations to charity, even when made anonymously, for example, fail to satisfactorily demonstrate altruistic motivations. Evolutionary game-theory is an increasingly useful field for understanding moral and political behavior, including altruism, and compliance to social norms. But contributions to this field, such as those by Kenneth Binmore, Robert Axelrod, and Brian Skyrms, have only successfully modeled cooperation, fairness, or altruistic punishment. They have not been able to establish the existence of altruism—they are only “how possible” and not “why necessary” explanations.

The existing empirical research on altruism can be said to fall into two main camps, neither of which adequately addresses the concerns about establishing altruism. In the first camp there are the experimental studies, such as those in game theory and behavioral psychology. Experimental or behavioral game-theory in economics, political science, anthropology, and psychology has attempted to capture emotional rather than rational motivations, but there are two major shortcomings to this research. First, game-theory experiments have mainly investigated how fair or trusting participants are in ‘dictator’ or ‘ultimatum’ games. Fairness and trust are

---


29 See Alexander Rosenberg for a good discussion of the difference.

30 The Dictator game examines how a subject will divide a resource between himself and another subject in a one-shot context. In the Ultimatum game, one person (proposer) is given the choice of
important ethical notions that may involve the emotions, but they are not identical to altruism. In addition, there is some evidence suggesting that subjects respond to the unconscious cues of testing environments and instructions. For instance, when subjects are instructed to “divide” a resource, this may bias a subject towards sharing a resource equally or fairly since “divide” seems to suggest an equal split. Thus some of the ‘fairness’ results can be called into question.

The next, and perhaps even more important, problem with experimental game studies is that they mainly focus on behavior and thus offer us thin descriptions of actions, whereby individuals are said to act in accordance with certain theories or act in a manner consistent with having particular motivations. For instance, the seminal work by Eckel & Grossman investigates contexts in which subjects make donations to others, but these studies do not give us thick descriptions of actions—i.e., those that dividing a resource between herself and the second person (responder), who then has one chance to either accept the terms of the division as they have been given to her or else refuse, in which case she receives nothing at all. If the total resource is, say, $10, and only divisible in dollar units, then if the proposer is “fair”, she will offer $5 to the responder and keep $5 for herself. Standard rational choice theory, however, predicts that the proposer will keep $9 for herself and offer only $1 to the responder, because if the responder is rational she would prefer $1 to zero. Since the proposer knows that the responder is rational and is also rational herself, then she will divide the total to keep most for herself.

Many researchers view the fairness outcome (50-50 split) as altruistic. The belief is that when a proposer divides a resource fairly, this division is altruistic because he could have kept all of the resource for himself instead—thus even though the proposer does not divide the resource so that the responder gets more than she does, a 50-50 split is still worse than a proposer could have done in an Ultimatum game, as predicted by rational choice theory. The problem with seeing a 50-50 split as altruistic is that there is good reason to predict that the responder will reject an offer if it is not fair. People get angry at individuals who have “poor manners” (Camerer & Thaler, 1995). Thus the proposer might just be pragmatically trying to avoid a no-win situation. Supporting this argument is the fact that in the “Dictator” games, where only the proposer’s decision determines the division of a resource and where participants are anonymous, by far the most common result is for the proposer to offer nothing to the responder (Burnham, 2003). So this seems to strongly suggest that it is a concern with the threat of rejection, rather than altruism or even fairness, that is leading the proposer to offer 50-50 splits in many ultimatum experiments.


For results and discussions see especially Guth & Tietz, 1990; Forsythe et. al., 1994; Hoffman et. al., 1994, 1996; Eckel & Grossman, 1996; Burnham, 2003; and Henrich et. al., 2005.
would actually investigate individuals’ motivations. The main exception to this work is that of Ernst Fehr and his colleagues who have attempted to find the neural correlates involved with punishing free-riders when punishers incur a monetary cost. However, even studies by Fehr and his colleagues on so-called ‘altruistic punishment’ have examined a very restrictive sense of altruism, if one even calls it altruism—altruistic punishment in their study could be interpreted to involve spite which does not seem to be the same thing as altruism.

In the second camp, there are psychological experiments, especially by Batson and Shaw, designed specifically for testing motivational altruism in contexts in which subjects are in apparent distress. Their valuable work shows us that it is implausible to believe that very crude theories of self-interest, such as the pursuit of material gain, can explain behavior in many circumstances. But critics argue that these existing studies on altruism do not adequately rule out more sophisticated forms of egoism—reductive altruism—and therefore do not adequately test for altruism. This is because (a) apparently altruistic motivations are compatible with more ultimate self-interested explanations (e.g., seeking internal rewards), and (b) even subjects’ self-reports of their altruistic preferences cannot rule out unconscious self-

interest. More importantly, many researchers believe that we cannot determine peoples’ ultimate preferences and goals. Therefore, they believe that there is no objective, scientific method of studying altruistic motivation and that we will forever be forced to rely on self-reports that can be misleading, erroneous, or reduced to self-interested explanations.³⁷

Researchers may be correct in their pessimistic assessment that there are significant obstacles to researching motivational altruism. But, importantly, they are only correct if one accepts the standard contemporary view of genuine altruism as being a function of ultimate preferences. It does seem plausible to argue that people’s ultimate preferences are not something that we have adequate scientific access to. Recall that researchers maintain that ultimate preferences can be unconsciously held—the individual herself may not be aware of them.³⁸ Furthermore, it is unlikely that we can discover the psycho-physiological or neural correlates to preferences, aside from whatever emotion or feeling component might occur—but then feelings or emotions become central, and this is not the standard way of viewing preferences. If we considered preferences as necessarily being affective, as involving feelings, as I argued in the first Appendix, many of the ways in which we explain or make sense of someone’s actions would not qualify as preferences. For instance, I currently want to


³⁸ A great deal has been written about the nature of desire in Hume. Michael Smith interprets Hume as maintaining that only some desires are experienced as feelings. More broadly, even aside from his reading of Hume, Smith makes a compelling case for the implausibility of viewing all desires as having a phenomenological feel. Smith, pp.92-10.1
continue typing, but it would be misleading to think of this preference in terms of an emotion or as being triggered by an affective state.\(^{39}\) This consideration also explains why first-person subjective reports are not viewed as adequate measures of peoples’ actual preferences. Since preferences are not necessarily affective, they do not necessarily all have a phenomenological feel. Thus it is reasonable to hold that people are not always aware of their preferences. Even if one were to hold that many preferences are affective states or involve affective states, however, the important point is that for the standard view of genuine altruism, altruistic preferences are not essentially about affective states.

A significant advantage of the feeling view of genuine altruism over the standard definition is that it may allow for greater testing for the conditions and existence of genuine altruism. For instance, work distinguishing the neural and physical correlates of emotions is quickly emerging.\(^{40}\) There is extensive work supporting different physiological response patterns between empathy and sympathy

\(^{39}\) Also, consider a long-term preference, for instance, the long-term preference to be a doctor. In moments of reaffirmation of this preference, following periods of doubt, the preference to practice medicine might express itself with a kind of force and perhaps even a feeling. But even in the absence of this feeling, during times in which the desire is not being challenged or reassessed, it would be odd to say that this preference simply goes away. It would also be odd to say that people no longer have preferences or cease to have them when they are sleeping.

on the one hand and personal distress on the other. To take just one difference, sympathy is robustly associated with lowered heart rate, and personal distress with increased heart rate. Furthermore, as discussed before, questions about motives are difficult to gauge through self-reports. But the feelings-view would seem to privilege first-personal reports. This is because while someone could plausibly have unconscious ultimate desires lurking in the background, it is far less plausible that he has unconscious feelings. While researchers maintain that we can be mistaken about or unaware of our ultimate desires even after sincere reflection and under normal circumstances, this seems far less true when we consider our feelings precisely because, as argued before, they do have a phenomenological feel. Thus, the feelings view of genuine altruism would seem to offer a potentially important avenue of testing for altruism.

---


42 One final difference between the two views of altruism is the following: whereas the preference-view maintains that there is one and only one kind of genuine altruism—the individual must have an ultimate desire for others’ welfare—the feelings-view is consistent with the recognition that there may be different feelings that are relevant to altruism depending upon the culture. After examining various cross-cultural research, Paul Eckman has famously argued that there are primary or basic feelings, such as anger and sadness, that are universal, and more sophisticated feelings, which are rooted in the basic ones, that vary across cultures. It may be the case that the feelings involved with altruism are of the latter sort. If this is true, then in order to study altruism in political life, we would first need to have an understanding of which feelings are central to a particular culture and then inquire into the prevalence of those feelings and the contexts in which they arise. (For instance, see Eckman, Paul. 1997. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System. New York, NY: Oxford University Press)
Part II Rational Choice and Consequentialism

3 The Structure of Moral Decision-making: Appreciating All of our Particular Ends

3.1 Introduction

This part of the dissertation turns to moral theory to focus on another feature of rational choice theory: its emphasis on consequences as the locus for determining whether a particular action, policy or institution is good or bad. As discussed in the Introduction to the dissertation, there are two ways in which consequences feature in rational choice theory. First, consequences, and in particular wellbeing measured by some form of utility or preference-satisfaction, explicitly figure in welfare economics. Welfare economics is the branch of rational choice theory dedicated to understanding what economic policies or actions make individuals better off. Importantly, whether or not someone is better off is determined by the relevant consequences. Second, and more subtly, consequences figure in the approach of rational choice. The rational choice theorist believes that to ask whether an action or policy is a good one is unhelpful unless we know what the alternative to that action or policy is. Rational choice is the process of choosing the best, however best is to be defined, selection among available options. Furthermore, the rational choice theorist believes that to ask whether one action or policy is better than another is already to imply a common measurement from which comparisons can be made. By being comparative in analysis, rational choice then places the emphasis on consequences, and specifically, on general consequences, because they serve as the basis from which to make
comparisons between competing options. In both of these ways, rational choice theory is a form of consequentialism and is therefore subject to the sorts of objections typically raised against consequentialism as a moral theory.

Many contemporary moral philosophers oppose consequentialism because, it is argued, it privileges abstract or general values over particular ones. For instance, the only ends that are valued for their own sake in consequentialist theories are the relatively general ones such as happiness and welfare. However, it is widely believed that, contrary to consequentialism, it is particular ends such as honesty and family and not general ones that are to be valued for their own sake. Related to this position on the proper source of value is the view that consequentialism is a psychologically demanding theory that alienates us from our own personal relationships and plans. More recently, and in a context largely unrelated to the problems of consequentialism, Michael Smith argues that the morally good person is someone who is motivated directly by her particular moral ends even when they are impartial or impersonal ends. It is this last criticism that I will focus on in this chapter.

---


2The second reason concerns a supposed disharmony between one’s motivations for acting, and one’s justifications for acting. The only justification for acting under consequentialism is to maximize good consequences, however defined. But this reason for acting is a very demanding one and might be very different from the less general concerns a person actually has and from the particular ends and virtues that a morally good person ought to be motivated by. The result of such a psychological and moral disharmony can be a diminished moral life.

There have been several attempts among consequentialist theorists to create what I call ‘two-level’ theories in order to address such concerns. In two-level theories there is a significant wedge between a justificatory general level of value, on the one hand, and a motivational particular level of ends, on the other. While there is a range of these sorts of theories and not all of them can be characterized as having different ‘levels’, I use this term only to convey the central point that general values and concerns are held to be independent of and take priority over particular ones in important ways. Defenses of such two-level theories, however, do not succeed in being full responses to the kinds of charges discussed above. While they seem to successfully address the problem of demandingness and the problem of psychological tensions, even the most sophisticated defense of such a theory—Peter Railton’s—does not yet adequately or fully deal with the concerns about (1) the proper source of our values; and, (2) the characterization of the good person. I address mainly (2) here, although I briefly discuss how (1) and (2) are related concerns and thus require a similar response.

After exploring the appropriate description of the good person, however, I argue that not only are two-level theories, and in particular, the motivational structures they contain, consistent with the appropriate view of the morally good person, but they are critical to it. This is because they are important for trying to reflect on our particular ends and for fully understanding and responding to the significance of all of our particular ends when we have more than one. Rather than a defense of consequentialism, and hence rational choice, based on its purported coherence or rationality, in this chapter I offer a defense of the kind of
consequentialism embraced by Railton based on what it means to care for one’s moral ends. Whereas Railton seems to argue that two-level consequentialists can be committed to their ends despite their consequentialism, I hope to show that it is because of the structure contained in these forms of consequentialism that consequentialists can be committed to their particular ends.

3.2 Against General Value

In simple consequentialist theories, the only things that are said to be intrinsically valuable—that is, valued for their own sake and not for the sake of other things—are relatively general things such as happiness, welfare, preference-satisfaction, and the like. However many scholars, especially in the virtue and care ethics traditions, argue that it is those things which are not so general or abstract that are intrinsically valuable. It is particular virtues and character traits, such as honesty, benevolence, loyalty and treating people with respect and equality, and our partial ends such as our families, relationships, and life projects that are intrinsically valuable. Bernard Williams, for instance, argues that we are alienated from our deep

---

While it is possible that particularity and partiality do not amount to the same thing—one could hold that intrinsic value lies in specific yet non-partial virtues such as treating humanity with respect and dignity (Philippa Foot seems to maintain this possibility in “Utilitarianism and the Virtues”—the two claims often go hand in hand. The proper objects of particular virtues such as honesty and loyalty are often those with whom we have special relationships, and indeed, with some virtues such as integrity, the object is oneself and perhaps nothing is more partial. Thus I will not distinguish here between literature that emphasize the value in particular virtues and literature that is concerned with the value in partiality. Instead I will lump both particularity and partiality under the heading of “particular” except for when the differences are important. (By ‘particular’ I am not referring to the way particular is used in the debate between moral theory and moral particularism). But because I will confine my discussion to arguments about ends, I will not include agent-relativity of the sort discussed by Thomas Nagel where reasons for acting in particular sorts of ways towards others can be said to be relative to particular individuals. (See ‘Autonomy and Deontology’ in *Consequentialism and its Critics.*)
projects and loved ones when we are motivated by general moral ends.\textsuperscript{5} Williams’ example of a man forced to choose between saving the life of his wife or a stranger illuminates his argument. If the man is moved by general and impartial values, then his motivating thought to save his wife might be something like “it is my wife, and in situations like this it’s permissible to save one’s wife”.\textsuperscript{6} But Williams famously objects that this man has “one thought too many”. Instead, the man’s motivation to save his wife should be out of direct concern and love for his wife; because she is his wife, period. Likewise, Susan Wolf argues that theories such as utilitarianism force us to strive for a kind of moral perfection.\textsuperscript{7} According to Wolf, utilitarianism is too demanding on us given that our normal psychology includes being motivated by desires and projects that are sometimes in tension with morality. More to the point, however, even if moral perfection were not so demanding—consider one who derives happiness from always helping and being kind to others—someone who attained moral perfection would be someone we consider stilted, humorless, and one-dimensional. We would not think it good to encourage anyone to be this way. Wolf’s argument seems to apply more broadly to the valuing of general ends when they are thought to be separate from particular ones. For it seems that the generalness or abstractness of the end of happiness in utilitarianism is in large part what contributes to it being a demanding theory. A view that would specify particular commitments to family and kindness only for those nearest and dearest would necessarily be more restrictive. Although this is not necessarily the case—consider a


\textsuperscript{6} Ibid., p. 101

consequentialist theory that specifies only one fairly undemanding obligation such as giving away $10 a year—the generalness of consequentialist ends seems to go hand in hand with its demandingness.

To address these challenges about general values, some adherents of consequentialism have introduced complex forms of the theory which contain ‘two-level’ structures, in which there is an ultimate level of value and a lived level of ends and traits. The ultimate level consists of some relatively general consequentialist values such as total welfare or happiness, and the lived level consists of particular practices, character traits, and ends that receive their justification through their contribution to the good consequences. While there is a range of these sorts of theories and not all of them can be characterized as having different ‘levels’, I use this term only to convey the central point that general values are held to be independent of and take priority over particular ends in important ways, as will be shown below. In addition, it is possible to have more than two levels of values and ends, but this would not affect the main issues in this chapter.

One two-level theory notable for its appreciation of the psychological importance of our particular ends is Peter Railton’s. Railton argues that since simple forms of consequentialism only regard good consequences, however defined, as intrinsically valuable, they fail to take seriously other, more particular ends and this detaches people from meaningful aspects of life. He proposes ‘sophisticated consequentialism’ in which an agent acknowledges the value of promoting good outcomes but at the same time acts by cultivating certain habits and character traits

---

8 Railton, “Alienation, Consequentialism, and the Demands of Morality”.

120
that enable particular commitments.\footnote{Railton, p. 109} A sophisticated consequentialist need not always be consciously motivated by the end of good consequences and so may do various things for their own sake. Because he does not need to bring consequentialist deliberation to bear on every action, he consciously acts for the sake of his particular commitments and so he does not suffer the fate of alienation.\footnote{Ibid., see especially p. 11 for a full explanation} Importantly, Railton demonstrates how his sophisticated consequentialism avoids collapsing into simple act-consequentialism.\footnote{Railton argues that unlike rule- and trait- consequentialism, sophisticated consequentialism is based on specific actions. A sophisticated consequentialist acts so as to maximize the good, but does not bring consequentialist deliberation to bear on each act. Because it is based on acts, he argues, sophisticated consequentialism does not collapse into standard act-consequentialism. pp.113-119.}

It does seem that Railton’s two-level view would reduce the psychological tensions problematic to simple consequentialist theories. If I am not always required to conscientiously deliberate on good consequences in order to bring them about, I can focus my attention on my family, friends, and the projects I hold dear. This will be less alienating—in my every day life I will be consciously motivated by my particular commitments and will thus be able to achieve a greater sense of emotional fulfillment through my engagement with them. This will also be less demanding in terms of my time and resources—I will focus on cultivating character traits and habits that motivate me to attend to those with whom I have special relationships. I will not need to constantly spend my time striving to help everyone in the world. Two-level views, with their mandates to focus on particular ends, do seem to be a great improvement over simple forms of consequentialism in terms of enabling an agent to live a rich and varied life.
However, two-level views come up against another sort of problem raised most recently by Michael Smith that would seem to be damaging to consequentialism as a *moral theory*—that they do not contain the proper description of the morally good person. While William’s concern, mentioned above, applies to the treatment of our personal ends, we will see that Smith’s concern is with the proper treatment of even impartial or impersonal ends and is therefore potentially more damaging to consequentialism. For even if one took a strong Williams-like point of view and insisted that only impersonal and impartial ends fall within the domain of moral *theory*, Smith’s concern would suggest that the consequentialist would still lack the appropriate commitment to his particular (moral) ends.\(^\text{12}\)

Smith begins with the claim that the good and strong-willed person (hereafter, the good person) is someone who is reliably concerned for or motivated by his particular value judgments. Smith’s goal is to argue against motive-externalism, which maintains that one can judge a course of action to be right and not be motivated to take that action.\(^\text{13}\) However his analysis also presents an apparent problem for consequentialism, even in its two-level form. In arguing against views that maintain a distinction between judgment and motivation, he begins with an example:

Suppose I am engaged in an argument with you about an ultimate moral question; a question about, say, whether we should vote for the libertarian party at some election as opposed to the social democrats...I come to the argument already judging that we should vote for the libertarians, and already motivated to do so as well...let’s suppose you convince me that I am...
ultimately wrong. I should vote for the social democrats, and not just because
the social democrats will better promote the values that I thought would be
promoted by the libertarians, but rather because the values I thought should
and would be promoted by libertarians are themselves ultimately
mistaken…what happens to my motivations?\footnote{Smith, p. 71}

Assuming fictional Smith (S) is a good person, we need to say how S acquires
a new motivation following this new particular judgment—that social democratic
principles are the right values. Unless we can say that his new particular judgment
itself causes him to be motivated, we will not be able to say that S is directly
motivated by his particular values.\footnote{Again, for the externalist, judgment is separate from motivation. So Smith’s motivation to vote for
the libertarian party, on the externalist account, is not a necessary consequence or part of judging that
libertarian principles are right. Therefore, when he comes to have a new judgment, now the judgment
that social democratic values are right, he may still have the motivation for libertarian values, since
judgment and motivation are separate. In such a case we would say that he is motivated to do
something he believes is wrong, which is not an accurate characterization of the good person. Thus,
we have the problem of explaining how it is that Smith, who is the good person, comes to have a new
motivation for social democratic values. The externalist cannot solve the problem merely by saying
that Smith’s motivation to vote for the libertarian party is changed to a motivation to vote for the social
democrats following his new judgment. For if Smith’s motivation to vote for the libertarian party is
due to a direct concern to vote for the libertarian party, then \textit{this} motivation cannot account for the new
motivation to vote for the social democrats.}

In order to maintain that there is any connection
between his newly formed particular value and a motivation for this value, we would
have to say that S has a more ultimate underlying concern which provides his
motivation.\footnote{Even though Smith discusses judgments of rightness or goodness and there may be important
differences between judging that \textit{x} is right or good and valuing \textit{x}, the differences will not be important
for what follows. What matters in this chapter is the structure of concern.} Smith maintains that this underlying concern would have to be a self-
conscious motivation to do the right thing in the abstract.

One can easily object to Smith’s description of the content of the underlying
concern the good person would be said to have. For instance, it is not at all clear why
this concern would have to be a self-conscious one, rather than a deeply held, but
partly conscious or even unconscious motivation. More substantively, it is also not clear why the good person must have a concern to do the right thing. There are other candidates, it seems, for the content of the good person’s concern. Perhaps he is concerned with being good. Doing what is right might be a large part of what is involved with being a good person, but it is not ultimately what the good person is concerned with. There are two differences between this possible description of the motivational content of the good person and Smith’s characterization. The first difference amounts to the distinction between someone who is concerned with taking actions and someone who is concerned with being a certain type of person. A concern to be a certain type of person does not always translate into discrete actions and might largely consist of having appropriate responses, emotions and character traits. A second difference involves the distinction between being motivated by rightness and being motivated by goodness. While being right does not seem to admit of vagueness or degrees, being good does. Being good seems to allow for the possibility of striving or attempting to be or do more, of being better or worse, but without necessarily reaching a determinate end. Someone ultimately motivated to be good might simply be someone who is driven to be better, even if only a little bit better—the psychology of such a person is such that it seems to involve an inherent understanding of one’s limitations and it suggests someone who is not relentlessly driven by moral notions. On the other hand, the picture of someone who is ultimately motivated to be or do right is of one who has a drive for a certain kind of moral exactitude or meticulousness.
Since we can simply take issue with the exact content that Smith has identified, it is more plausible to interpret Smith’s objection to be targeted not at the content but at the *structure* of concern—the fact that the good person’s concern for his particular ends would be ‘indirect’ and ‘mediated’ by other, more ultimate and general values. Thus the important implication of Smith’s analysis is that S’s concern for socially democratic values is indirect, since it is provided by his more ultimate concern.

But Smith takes this to be a problem. He argues that the good person cannot be said to be concerned indirectly for the particular things he judges right to do. Just as good lovers must have a direct, unmediated concern for their loved ones, good people, he argues, must have a direct, unmediated concern for equality, justice, honesty, the welfare of people, and other particular things that they believe are right. He writes that “a view that holds that the good person cares only indirectly for these things and cares directly only for doing the right thing, is absurd: it elevates a moral fetish into the one and only moral virtue”.17 By this last comment, it might seem as if Smith’s argument merely reiterates Wolf’s argument against moral perfection—that a morally good person would end up being one whose personality is overwhelmingly characterized by moral concerns and as such would lack the complexity and richness of a normal, healthy personality. But, importantly, the good person in Smith’s discussion of externalism need not always be acting for the sake of morality—he need not experience the psychological and effort demands caused by adhering to simple consequentialism. Rather Smith’s worry about externalism can be stated as: *insofar as*

---

17 Smith., p. 76
or when one considers the particular ends that are morally valuable, his concern for these particular ends is indirect. The only direct concern that the good person would have under externalism is doing what is right in the abstract. The problem with this description of the good person for Smith is about the indirectness of his particular concerns and not about the extent of them or the fact that all of his concerns, including appropriately non-moral ones, would be subsumed under some general moral value.¹⁸

What this would seem to mean is that even two-level consequentialist views do not include the proper description of the good person. For even if we take the ultimate concern to be promoting good consequences, rather than doing the right thing, and we maintain as Railton does that we can frequently consciously act directly for the sake of our particular ends, the seeming problem for two-level views is that they do not contain the appropriate structure of concern, as I will show in greater detail in the next chapter. Even sophisticated consequentialists must mediate their concern for their particular ends by appealing to more general values, and this is especially evident in cases of conflict. However, I will argue next that this indirect concern is not obviously a problem, as Smith and other authors would have us think. On the contrary, the structure of concern contained in two-level views is, on reflection, precisely the sort we would want the morally good person to have.

In the sections below, I do not reject two-level views but instead attempt to provide a defense of them in response to the charge about the morally good person. Since Railton tried to defend sophisticated consequentialism by arguing that an

¹⁸ Wolf’s central claim can be summarized as following: her arguments “call into question the assumption that it is always better to be morally better.” P. 438
adherent can value particular ends for their own sake, the implication of his argument is that it is *despite* the two-level structure of concern contained in such views that sophisticated consequentialists can be committed to their ends. I hope to show that it is because of the structure of concern contained in two-level views that consequentialists can be committed to their particular ends.

### 3.3 Mediating Ends in Our Everyday Lives

Before I proceed, I want to emphasize that problems of incommensurability between values loom large, and in what follows I do not assume any particular kind of commensurability between our ends, and especially, any objective commensurability between our ends as is sometimes assumed in rational choice. My analysis instead proceeds from the observation that we do engage in, and must engage in, comparisons between our ends in many situations. Given this fact, my concern is how we should consider the comparison of ends and in particular, the comparison of ends by the morally good person. I will not argue for a specific standard of comparison or attempt to demonstrate how any standard can resolve the problem of comparing our ends. Rather, I hope to show here that many of us already implicitly rely on something like two-level views in *trying* to compare ends, even if no easy answer can be found, and it is this trying that is important. Importantly, though, when we do try to address conflict between ends we both appeal to more

---

19 Even if we cannot address all of the problems of commensurability, there are a few considerations on this topic that support the arguments here. First, at least some of the work on this topic suggests that even if our abstract values are not commensurable, our particular ends are at least comparable. Furthermore, even in the absence of true comparative evaluative judgment, this would not seem to defeat the claim that we can still *try* to compare our ends, and it is this claim that is central to the analysis in this chapter. See Elijah Milgram (1997) and Henry Richardson (1994)
general values, and treat these values as being more ultimate in certain ways than the particular ends we are trying to adjudicate.

A discussion by Dworkin demonstrates the problem of comparing our ends. After claiming that we appreciate the value of art for its own sake, Dworkin asks, with respect to the value of human life, that “if it is a horrible desecration to destroy a painting, for example, even though a painting is not a person, why should it not be a much greater desecration to destroy something whose intrinsic value may be vastly greater?” Dworkin makes two claims here: (a) that we treat both great art and the life of persons as having intrinsic value in the ultimate value sense; and, (b) that we can compare their values. Dworkin suggests that art, and presumably persons, also have value because of their contribution to something else such as pleasure. But even though we could compare the values of art and persons in terms of their contribution to something else, it is not this sort of comparison that Dworkin is referring to when he suggests that the value of persons is vastly greater than the value of a painting. It is their relative intrinsic values that we are being asked to compare.

Many of us do have the attitude expressed in (a). A problem seems to arise though precisely because many of us have the attitude expressed in (b). How are we able to hold both (a) and (b)? For if we are committed to particular ends as valuable in themselves and not in terms of other values, by what means are we comparing them? Given that we do make such comparisons, does this mean that we are just

---

20 Dworkin, Ronald, “What is Sacred?” in Bioethics, ed., by John Harris. (New York: Oxford University Press, 2001) p. 158. He demonstrates the ubiquity of the attribution of intrinsic value by discussing our attitude towards great paintings. We believe that art has its own value apart from its contribution ultimately towards something else, be it our education, pleasure or wellbeing, or anything else. I would add that we furthermore seem to believe that any particular great artwork has value in itself, apart from its contribution ultimately towards the total category of great artwork.
choosing at random, merely tossing a coin between which of our ends matter more? Or do we have some recourse for guiding us through these decisions?

In difficult scenarios, we often appeal to indirect and general concerns to adjudicate between competing particular commitments, even very personal ones, and these occasions are much less rare than some authors seem to assume. If we take our work, our health, our families, and our favorite hobbies to be serious commitments, and it seems many of us do, then we are forced to choose between these ends very often. For our particular ends come into conflict frequently, even if we have more or less settled on how to resolve them. The fact that we do not consciously experience tension among our different projects and relationships or that the tension does not feel like an intrusion on our lives does not suggest that our projects and relationships do not conflict. Often we fail to realize that we are invoking common general values, but this is because we have adjudicated these sorts of conflicts many times before, and so solving them at any given time can be habitual. If I am at the library researching books for a new paper that I would enjoy writing and start to feel that I am coming down with a cold, I might quickly choose to stop working and go home to get rest. Even though there is no feeling of real conflict here between my health and love of my work, that does not mean of course that conflict is absent. We have all faced a number of decisions such as these in the past and have settled on rules of thumb for negotiating them—such as, “if I have the luxury of time to take care of myself when I am feeling bad, I should stop working and do so”. That there is genuine conflict between my health and my work is obvious in the fact that I would not always choose my health over my work. If I am at the library finishing a
conference paper that is due the next day when I begin to come down with a cold, I would not choose to stop working. The general value at stake in both cases is simply that I want a balance in my life between pursuing my work and my health. If I need to compromise one for the sake of the other, then it should be the one that is least pressing at that point. Many of us try to balance our disparate ends and internalize this balance in such a way that we are not constantly faced with the fact that we are required to reconcile our different engagements. And the balance is achieved not by appealing to one of our particular ends, but to more general values. Having to make a decision between two commitments when they conflict on a given occasion can either reveal more vividly what these general values are or become cause for determining them.

To see these points, consider the following sort of scenario. Bob has a deep and rewarding relationship with his wife and loves her dearly. His best friend, whom he has known for years, is also very important to him. Each of them, on a long weekend, wants to spend time with him. His wife has just been rewarded at work and would like to celebrate by taking a small trip, but his friend, who lives in the next state, has been depressed and calls Bob to request some company for the weekend. In deciding whom to spend the next several days with, Bob might try to appeal to his value of attending to the person whose need is greater, and decide to visit his friend. Not only would this relatively more general concern—attending to the person whose need is greater—mediate Bob’s decision not to spend time with his wife, but his action for the sake of his friend would also be mediated by appeal to this concern. Furthermore, when there is tension between Bob’s concern to attend to the person
whose need is greater and some other general concern, such as giving priority to his wife because of the deep love and respect he has for her, it seems he will need to try to mediate by a still more general concern than either of these. Bob might feel that always giving priority to his wife would not allow him to have other rewarding relationships and this is important not only for a full and varied life but also for the sake of a healthy marriage, which in turn enhances his life and enables him to better respect his beloved wife. Bob would be trying to mediate his particular ends in this case by attempting to find a general value to enable him to decide which particular end to give priority in which situation.

Again, this is not to suggest that more general values can satisfactorily help us decide between competing particular ends, but it does seem general values can do so better than particular ones can. In the example that I just gave, it is not clear how either of the particular ends—either Bob’s wife or his best friend—is helping him decide what to do. Nor does it seem likely that some third particular end could help. How could, for instance, spending time with his daughter, or on his hobby, help him decide to visit his friend? A relatively more general value can help decide precisely because it is less contentful—because it relates to both particular ends. The concern to focus on the loved one whose need is greater relates to Bob’s commitments to both his wife and best friend because they are both loved ones. What is left for Bob to determine is whose need is greater, but he will only do this once he has already invoked a general concern that relates to both ends. This last point about general values relating to particular ends gives rise to the more important reason, in terms of the goal of this chapter, for appealing to more ultimate general values in deciding
conflict. This reason is that it enables us to take all of our ends seriously and is the subject of the next section.

It could be argued that I have not demonstrated the importance of two \textit{levels} of concern, but only that we need two different kinds or categories of concern. My examples may only have demonstrated the need to have both general and particular concerns, with the former having to frequently adjudicate the latter. This kind of position would not necessarily present a problem for my arguments, since the claim is that there must be some stable division between general values and particular ends, as is the case with two-level consequentialist theories, and it does not matter for the central concern here precisely how we characterize this division. However, in all such cases where there is a division between general values and particular ends, it will turn out, it seems, that general values must be in some sense or at a given time more ultimate than the particular ends if they are to settle conflict in our ends. To see this, consider the contrary view that would insist that we can bring other general considerations to help adjudicate conflict in our particular ends without requiring that these general considerations constitute more ultimate values. But then we must ask what the status of these other considerations is if they are not more ultimate values. For it seems that general concerns can only prioritize and adjudicate conflict in our particular ends if they have a certain priority themselves; a priority over the particular ones—by being at least in some regards more ultimate values. If the more general concerns did not have priority in this way, it would be unclear how or why we would observe or abide by them or why we would let them dictate what activities and relationships to give attention to, especially given the already assumed importance of
these particular ends. This does not suggest that general concerns must always be more ultimate than particular ends or that there must be a fixed order between general values and particular ends. Perhaps it is possible to maintain that a general concern is only held to be more ultimate than particular ones at the moment one is trying to decide conflict or reduce tension between particular ones. However even if this were the case, the charge against mediating our particular ends through more ultimate general values would still apply, if only for that given moment or time that one is relying on a general value to try to mediate particular ends. I will return to this topic of the status and ordering of general values in the final section.

I said that as long as there is a stable division between general values and particular ends, with general ones taking priority over particular ones in certain ways, it does not matter whether we characterize the account as containing levels or in some other way. However, there is one kind of account that would seem to be contrary to my arguments, and it is the view that certain particular ends are constitutive of more general values or that particular ends are instantiations of more general ones. Joseph Raz, for instance, seems to hold this type of view when he argues that the claim that something is capable of being good or valuable for something else is not equivalent to the claim that it is an instrumental value. He writes, “when one says that reading Proust enriches one’s life one is not pointing to the consequences of the reading. Rather reading Proust with understanding is such enrichment.” The problem with this view is that it does not seem to help us address the problem of having to trade-off competing particular commitments. If Proust is enrichment and so is Joyce, then

when I want to read both Proust and Joyce but have only time for one, how am I able to try to use enrichment as a guide for distinguishing between them? It seems necessary to hold that general concerns are *separate* from particular ends, that they are independent enough in order to serve as guides. In that case, when I am pressed for time, I might decide that Proust would be more enriching than reading Joyce or that since both are enriching in equally but different ways I will have to read one now and one later. I might simply toss a coin between them to decide which I delve into first, but even if I do this, it seems I have come to this conclusion of tossing a coin by reflecting on, even if only momentarily, a concern to engage in enriching activities, *apart* from my interest in either Proust or Joyce. I appeal to the general concern because I need a value by which to compare the particular ends, but it seems that the only way I can do this is if the general concern has a certain independence from the particular ends.

As with the previous example of Bob having to decide whether to spend the weekend with his wife or with his friend, general concerns mediate the more particular ones by providing standards by which to compare them, by telling us which particular end should receive our attention in a given situation and by revealing the way in which they should be ordered or prioritized. Again, though, it seems the more general concerns can only do all of this by having a certain priority themselves. But in order for a general concern to have priority, it must be held as being at least somewhat independent from the particular ones. The general concern would not be in a position to operate as an arbitrator of particular ends if particular ones were instantiations of it or if the particular ends constituted the general one. Even though a
general concern can be said to relate to particular ones and to reveal the shared relationships between them, it must, at the same time, be separate from particular ends.\textsuperscript{23}

In the beginning of this section, I presented Dworkin’s claims that we both believe our particular ends are valuable in the ultimate sense and that we can compare these values, and I suggested that these claims might be inconsistent. While there may be a way to reconcile them, I hope to show that we do not necessarily have to. Specifically, we can abandon the belief that particular ends are valuable in an ultimate sense while still being fully committed to them. As mentioned in the Introduction of this chapter, even though they are separate issues, a single response might serve in addressing both (1) the problem of locating the proper source of value of particular ends and (2) describing how the good person treats his particular ends.\textsuperscript{24} It seems we would want to describe the good person as treating his ends to be valuable in the sense that they are in fact valuable. This suggests that we can determine how the good person ought to treat his ends only by first determining the source of their value. Thus, if we hold, as so many authors do, that our particular ends are valuable in an ultimate sense, then the good person does not try to mediate them through more underlying general values, for there are no such values available. However I argued that many of us do try to mediate our ends through underlying general values. If we do see (1) and (2) as related, which seems appropriate, then an implication of my argument is to question whether it is the case that the good person’s particular ends

---

\textsuperscript{23} One comment related to this point about separateness is that Raz’s kind of view strikes me as insufficiently distinguishing our particular commitments, and therefore as not emphasizing the significance and uniqueness of our particular ends from each other and from general ones enough.

\textsuperscript{24} I discuss the issue of (1) in greater depth in the next chapter. In addition, I have not of course addressed which values should serve as ultimate ends. This is beyond the scope of this chapter.
are ultimate values. But is this necessarily cause for concern? Perhaps the most important reason for doing (1), attributing the proper source of our values, especially in the context of the worries expressed over abstract ethical theories, is to ensure that we take our particular ends seriously. The way in which we can say the two questions are related then would be to say that the good person must be someone who takes his particular ends seriously. What we need to determine, then, is what it means to do this.

In the next two sections, I will argue that two-level views do enable and indeed encourage us to treat our particular ends seriously—and this is by ensuring that we consider each of our particular ends and that we do so in a stable fashion. I will say much more on the topic of ultimate or intrinsic values in the next chapter.

3.4 Considering Each of Our Ends

As discussed above, I am not defending the claim that appealing to general values satisfactorily helps one solve dilemmas or conflicts in one’s particular ends, although as I argued before, I do claim that they do so better than particular ends could. My goal is to instead defend a broad range of views against the charge that in virtue of appealing to more ultimate general values for the sake of resolving conflict in one’s particular ends, whether one can successfully do so or not, one does not have the appropriate commitment to his particular moral ends. In fact, my claim is that it is by appealing to more ultimate general values, and not in spite of this as some consequentialists have suggested, that one’s commitments to his particular ends are deepened and gain stability.
To illustrate these points, I return to examples from personal contexts, in this section, and then examine more obviously moral contexts, in the following. Although I do not assume any division between personal ends and morality, I hope to show that even in the very personal contexts, two-level structures to decision-making encourage a deep commitment to one’s particular ends. There are two reasons for beginning with personal scenarios. First, many authors, such as Williams, are concerned with the nature of one’s commitment to personal ends in consequentialist theories and my analysis will suggest that charges such as alienation from our personal ends are, at the very least, too strong. Second, and more important for the direct goal of this chapter, if it can be shown that two-level decision making can encourage a robust commitment to one’s personal ends, then this point will be even clearer in the sorts of impartial scenarios Smith has raised in assessing the characterization of the morally good person—and the success of my arguments mainly depends on the latter.

Recall Bob who has to decide whether to spend the weekend with his best friend, who is depressed, or go away with his wife who would like to celebrate her recent success at work. Even if Bob is ultimately unsuccessful in resolving the tension between spending time with his friend and his wife, the very act of trying to compare them by searching for a common standard of value encourages him to: (a) appreciate his ends taken together; and more importantly, (b) consider each end. Taking the first point, the concern to focus on the loved one whose need is greater or the concern to have full and rich relationships with others reaffirms the fact that both Bob’s wife and his best are loved ones and that they each, in their own way, enhance his life, and thus also brings these separate particular ends into a relationship with one another.
Relating each particular end to each other through a general and more ultimate value seems to embed each end into a framework of things that are important to Bob, rather than each one standing alone. That Bob’s relationship with his friend is valuable because it is related to his general value of having a fulfilling life, with many loving relationships, connects his appreciation for his friend with his other significant relationships, including with his wife. It connects his commitment to spend time with his friend to the other things he cherishes. Rather than this seeming alienating or that he is not fully committed to his particular ends, relating these ends to a general value and so also to each other connects them together and may bring out their combined importance. When we have many particular ends that are very significant to us, and of course many of us do, the appeal to general values can be seen as the recognition that each and every one of our particular ends is special, unique, and important and that we have many more than one such end. Mediating particular ends through more ultimate general values enables us to recognize the totality of our ends, and in this way, our ends are being treated seriously.

On the other hand, trying to resolve a conflict between ends without any more general values to serve as common standards would seem to only give rise to meditating on each particular end in isolation. By treating his commitment to his wife and to his friend as separate and ultimate, Bob may fail to see how they are related and therefore fail to see other ends or fail to fully consider other ends that are not being attended to whenever he decides to favor any given one. This point brings me to (b), the second and more important reason for the importance of trying to mediate
conflict between ends through more ultimate general values—and that is to consider and appreciate each end.

When critics charge that the appeal to general values causes alienation from a particular end, they seem to be implicitly making this claim from the point of view of the end at hand, or the end one is choosing to attend to. For instance, Bob’s decision to visit his friend by appealing to his deeper value of attending to the person in greater need, and the further value of having a fulfilling life might leave the critic cold. It might seem that Bob’s need to justify spending time with his friend by some other value—other than his friend—diminishes the significance of that relationship, and indeed this has been the charge. However, what such a critic is failing to focus on is the forsaken end; the end one has decided not to attend to at that time. Bob’s forsaken end in the example is his relationship with his wife.

If Bob were to choose to visit his friend simply because it is his friend, then when his wife asks Bob why he has chosen to spend time with his friend instead of with her, his response would be that he chose this way because it is his friend. Consider though how this would make his wife feel. Could she not reasonably assert that she is his wife and that Bob should feel that the fact she is his wife is just as good of a reason, if not more, to go away with her for the weekend? If Bob were to directly appeal to his commitment to his friend then this would seem to leave his wife feeling that she is not as important to him as is his friend. Furthermore, she would not be unjustified in thinking that Bob does not seem like he has given her much thought in this decision. On the other hand if, as in the example, Bob tries to mediate both of his commitments through his deeper values and concerns, then his response to
his wife will be very different. He would say that he has ultimately decided to spend
the weekend with his friend because his friend seems to need him more at the time
than she does. Furthermore, he would say that even though she is his wife, he cannot
always prioritize her because it is important to have many deep relationships in his
life, which will in turn enhance the strength of their marriage and his commitment to
her.

I am not suggesting that the second response would placate Bob’s wife or
would leave her happy to spend the weekend alone. I am instead making the
comparative point that this sort of appeal is more likely to make his wife feel that Bob
values her, because it reflects the fact that Bob considered deeper values in making
his decision not to spend the weekend with his wife—that he did not only consider his
friend. Thus from the perspective of the forsaken end, appealing to more ultimate
general values in situations of conflict implies that the decision maker has, at some
point at least, reflected on and deliberated about both ends, including the forsaken
one, which in itself demonstrates his commitment to that end despite the fact that the
end was not favored in this scenario. Appealing to more ultimate general values, in
the way that Bob has, forces us to stop and examine the loss involved in not choosing
other ends when we decide to favor certain ones, and the importance of this is
difficult to overstate when we have many precious ends.

While a commitment to such a general concern as wanting enrichment might
be more ultimate than particular ends in some regards, this would not imply that one
is not committed to the more particular ends. Rather, appealing to a more general
ultimate concern can be seen as a means for trying to contemplate and balance our
many disparate particular commitments. I am reminded of a good scene from a mediocre movie, *Erin Brokovich*. In this scene, a character who is a lawyer tells the character Brokovich that she is taking the legal case on which she is working too personally. Brokovich emotionally declares, “This is my work! My sweat! My time away from my kids! If that’s not personal, I don’t know what is!” Brokovich understands that she has made trade-offs among her particular ends in order to live more fully. She works long hours and very hard in her job not only to earn more money to care for her kids, but also to feel that her ideas and skills are important and respected. By valuing a full life, she has traded-off some of her time with her kids for the sake of another particular end—the quality of her work. This does not imply, however, that she does not genuinely care about her kids. Indeed, it is mainly because she cares so much about her kids that she takes her work so personally—she needs to feel as though her time away from her kids is *worth* it.

We often feel as if our particular commitments, such as our families and projects, are ultimate concerns, with no other concerns lying beyond them, and certainly nothing general or abstract lying beyond them. We might even say that our particular ends are ultimate in certain areas or domains of our lives. In the domain of one’s day-to-day projects, one’s research might strongly override all other undertakings. But even our most worthwhile projects come into conflict with other domains, such as when doing research conflicts with one’s health or with duties to one’s family. We must often weigh and trade against one another those things that are most precious and special to us and do so by appealing to more ultimate and general concerns.
3.5 Mediating our Moral Values

As with our personal ends, many of our impartial ends, duties, and virtues come into conflict with one another frequently, and there will be a need to adjudicate between them. To say that they might conflict is not to require that they are somehow at odds or in opposition with each other. Indeed, Philippa Foot argues that separate virtues provide constraints and form to one another. The virtue of friendship, for example, plays its part “in determining the requirements of benevolence, e.g., by making it consistent with benevolence to give service to friends rather than to strangers or acquaintances.”25 So we need not think of particular ends and duties as being necessarily exclusive of one another. Beneficence need not be at odds with the duty of respecting others. But the actions required to fulfill one duty might leave another one unfulfilled, and for deciding between them, we need to resort to values independent of both. For instance, one might judge that it is right to give money to charities that help the poor in developing nations, but also believe it is right to support organizations that try to promote equality for women in Muslim nations. Deciding how much to give to each, or whether to give a donation to only one cause, would require one to consider other values and, again, it seems the only way these other values can adjudicate is if they have a certain priority themselves. As I argued earlier in the chapter, maintaining that particular ends either constitute or instantiate more general values would not seem to adequately address the problem of priority among values.

25 Foot, pp.235-36
If we insist, as Smith does, that the good person does not mediate his particular moral ends through other values, even from time to time, we will have trouble explaining how he attempts to deal with conflicting ends; how he decides in the face of conflicting particular concerns. We could of course say that the good person would then simply be directly motivated by both ends. But we will not be able to explain how one ends wins out, so to speak. On the other hand, if we describe the good person as caring about his particular ends but also as being committed to trying to mediate them via more general and ultimate ones, then we will be able to say that the good person resorts to these more ultimate concerns when he is faced with conflicting particular ends, and thus be able to explain how one or the other end, or a different one altogether, prevails.

A likely response to what I have just said is that the good person need not be so rational and practical in adjudicating his particular commitments. Even if it is practically important to hold that when there is conflict between particular ends one’s concern for them must be mediated by more general values, this does not seem to square with our intuitions about our typical, everyday moral concerns. I am greatly upset by harm to animals and it does not feel as if my actions on behalf of animals are motivated indirectly, mediated ultimately by some general concern. But when I consider the full picture of what would be involved with being directly concerned for the welfare of animals along with having a direct concern for each of my other particular moral ends, it seems more plausible that I relate these particular ends to a more general value, such as wanting to alleviate cruelty in the world, which may itself be related to an even more general value. Even aside from immediate conflict among
particular ends, it does not seem plausible that the good person would not have more general values as underlying ones because it is not plausible to think of him as having particular concerns that are not related to each other in some systematic and stable way.

To see this, consider again the fictional Smith (S). After discussing the two political positions with someone else, S no longer judges that voting for the libertarian party is the right thing to do. He now judges that social democratic values are right. But what has happened to his concern for libertarian values? If S is a good person, then by his own criteria, what he cares about when he changes his particular judgment is precisely what he judges it right to do at that time. But this implies that a moral concern can completely fall away and another one can take its place as a particular judgment changes. There are several reasons why this is important. First, it would not even make sense to say that his new particular concern, now in social democratic values, takes the place of his old concern, since there is no proper ‘place’ of which to speak. S’s new direct concern for social democratic values is not related to his, now absent, concern for libertarian values. For under the view that we are motivated directly by particular values without any more general concerns mediating them, it seems S’s concerns for the two different values are unrelated. But then, it seems to me, the question we must ask is how the normative debate in which S is challenged by his friend influences him to alter his judgment about which party is right to vote for. If his concern for libertarian values is direct, then there would seem to be no special reason why normative discussion about the virtues of social democratic values would be of significance to him and cause him to reject his
libertarian values any more than, say, a discussion about the virtues of wine—since
the virtues of wine would be no more or no less distinct from the virtues of libertarian
values than would be the virtues of social democratic values. Merely stating that the
two political positions are inconsistent whereas the virtues of wine and the virtues of
libertarian values are not inconsistent does not resolve the problem. We would want
to say that S’s concern for libertarian values and his concern for social democratic
values are related, that they stem from the same ultimate concern, in order to explain
how his new judgment and motivation come to supplant the old ones. We can
furthermore ask why S should care about whether his libertarian values are right in
the first place unless he has a concern to discover what political values are worth
having.

A second reason why it is important to note that S’s concern for libertarian
values simply ends as he changes his particular judgment is that his values are only as
stable as his particular judgments are. To see why this is a problem, consider the
contrasting view that is being defended here. With a two-level structure of concern,
general concerns would underlie more particular ones. So when S would acquire the
judgment that the social democrats are right, his new concern for social democratic
values would be provided by this new judgment together with, and mainly from, his
already existing, underlying values. A plausible way to think about this is to hold that
his change in particular judgment would refine and provide specific shape to his
existing, underlying general values. On the other hand, under Smith’s view, when a
particular judgment changes the good person would lose a concern altogether, instead
of just the resulting concern that is given its specific character by the judgment. But it
seems that this would be inconsistent with how we would want to describe the good person. We think of the good person as someone who reliably has certain deep and general concerns and values that transcend particular judgments, not merely a collection of direct concerns that are only as stable as are his particular judgments about what is right or good, or kind and compassionate. We do not think of him as having a series of unconnected direct motivations for particular things like honesty, equality, and the well-being of his children, without any broad, more coherent, and more stable way of being concerned for all these things together. More specifically, it seems that the good person is someone who has stable patterns of particular moral commitments and having these implies having a way in which a particular commitment is connected with and fits and works together with other particular commitments. Having more general ultimate values is important to achieve this, as the examples from above attempted to demonstrate. The good person need not be so rational as to have all the details worked out—indeed it seems we would want to encourage some flexibility in thinking—but it is important that he is willing to consider how particular ends are or might be related to one another.

This last comment brings me to my final reason why we do not think of the good person as one who has distinct and direct particular concerns. As I suggested above, in our personal lives we tend to balance our separate particular ends in a way in which we are not constantly faced with having to make decisions. Many of us reflect on and balance our disparate ends and internalize this balance in such a way that we are not constantly faced with the fact that we are required to reconcile our different engagements. But more important than someone who cares deeply about all
of his personal ends, the good person is someone who periodically reflects on and compares his particular moral ends, even apart from being faced with immediate conflict among ends. It might be very important to someone who cares deeply about his specific personal commitments to reflect on them and try to harmonize them as much as possible; it would probably enable him to better determine how to go about his life and help him achieve a greater sense of fulfillment. But one could argue that it is not constitutive of someone who cares deeply about his personal commitments to be willing to reflect on all of them. Your commitment to your spouse might trump all other particular personal commitments no matter what, and you might be unwilling to hold this commitment up to scrutiny. Similar to William’s example, this might be the best way to think about your spouse—your unwillingness to question your commitment in fact may indicate just how deeply you care about this particular value. Even though this view about our personal ends is debatable—consider someone’s unwillingness to scrutinize his marriage for fear of discovering that it is not a good one—this is not how we think of a morally good person. It is constitutive of the good person to be willing to reflect on and compare his particular moral ends from time to time, to assign importance to them and consider how much each matters, even if only to ponder how hopelessly difficult it is to prioritize his particular ends. But if he is doing this, then at least implicitly, he is trying to rely on more general ultimate values.
3.6 Conclusion

It seems clear that simple consequentialist theories do not reflect the good person’s dispositions towards his particular ends—it is important for reasons made clear by Railton, Williams, Wolf and Smith that the good person is not always focused on general values and, especially, is not always thinking of general moral values. I have also argued, however, that treating our particular ends as direct and unmediated is not consistent with the description of the good person either. The good person must be someone who takes his particular ends seriously and two-level views do enable and indeed encourage this by encouraging us to consider each of our particular ends. Without trying to mediate our particular concerns through more general underlying ones, we would be treating the former as separate and ultimate, and the problem with this approach is that it can cause us to overlook what other concerns are being forsaken whenever we abide by or attend to any given one.

Before ending this chapter, is important to address a potential misunderstanding of the arguments presented here. This can be brought to light by raising another of Williams’ objections to systematic moral theory and to consequentialism in particular. Williams argues that it is crucial to avoid sparseness in our ethical resources—to avoid reducing all of our ethical notions to just one, as is often explicitly done in rational choice theory.26 My arguments in this chapter do not attempt to deny this. Appealing to a more general common value to try to mediate tension between particular ends does not mean that we must reduce them to a single

26 Williams, Bernard. 1985: Ethics and the Limits of Philosophy. London: Fontana, “Our major problem now is actually that we have not too many but too few, and we need to cherish as many as we can” (p. 117)
value. Furthermore, I am also not suggesting that we will not face real dilemmas between our particular ends, for there surely will be tragic conflicts.

I have argued that unless we compare and reflect on our particular ends by trying to mediate them through more general values, we run the risk of neglecting some of our most precious ends. However, it is not necessary for the purpose of mediating and comparing our particular ends to relate all of them to a single ultimate value. For, importantly, it is not clear that we are merely pushing the problem of having to compare our concerns back to the more ultimate level. As I have argued, our more ultimate values must be somewhat general if they are to serve as mediating and resolving tensions between other concerns. But since they are more general, it seems they will not come into conflict with each other as frequently or as sharply as our particular commitments do because, among other reasons, they will not enter our lives or our societal or institutional practices nearly as regularly. We will frequently be in the position of having to trade-off particular ends, such as protecting the habitats of certain animals and promoting the needs of laborers in developing nations. On the other hand, the values of compassion and of equal respect will less frequently conflict. But there is a more subtle and important reason why two-level views may not merely push the problem of having to adjudicate our values back, and this is because it might not be as morally important to frequently determine how general values conflict. Because our particular ends are so significant and serious to us, it seems more important to frequently consider each of them and the relations between them, and to take general values as fixed foundations for mediating and reflecting on our particular ones. Consider an imperfect analogy from architecture. It is critical to
have a strong foundation to a house. But once a foundation is secure, it seems more important to attend to the roof, windows, and doors—the very stuff of the house, the parts of the house that are most significant for our everyday living. On any given occasion, failing to consider the foundation is unlikely to have dire consequences even when there are problems with it, whereas failing to consider a broken window could be disastrous. From this last point it could be suggested that precisely because our particular ends are so important for our lives, we would not want them to form the foundation of our values.
Part II  Rational Choice and Consequentialism

4  Cost-Benefit Analysis in Ethical Decision-making

4.1 Introduction

In the previous chapter I presented the charge that consequentialists who do not place intrinsic value on their individual projects and relationships are alienated from meaningful aspects of life. In this chapter I will discuss how similar sorts of charges have explicitly been made in the context of rational choice theory—and this is against the use of cost-benefit analysis in moral decision-making. Cost-benefit analysis does not always assume that the ends or values to be compared are consequences, but does possess the mode of ethical thinking I described above. In this chapter, I will argue that while traditional cost-benefit analysis that relies on some notion of efficiency, such as pareto efficiency, is subject to a wide range of moral objections, cost-benefit analysis as conceived as a framework in which to compare our moral ends encourages broad reflection and the consideration of each alternative value. David Schmidtz has made a similar argument regarding cost-benefit analysis, however my position diverges from his in significant ways. Furthermore, I will argue that far from calling for the maximization of any value, cost-benefit analysis simply calls for the minimization of costs, or more minimally, the assessment of costs—and costs are not strictly monetary ones but are rather the values and ends which are not favored whenever an action is undertaken for the sake of some other end or value.
The context for my discussion on the use of cost-benefit analysis in ethics is bioethics. Since many bioethical positions and policies must weigh the value of life or particular lives against other, sometimes more general values, they can be described as failing to treat life as always intrinsically valuable. So it is especially fruitful to provide examples and illustrations from bioethics to motivate the discussion. I argue that far from causing estrangement from our significant values, the broad sort of cost-benefit analysis advocated here allows us to consider all of our important values and, in this way, take our values seriously.

4.2 Intrinsic Value and Alienation

The area of bioethics is laden with claims about the intrinsic value of life, treating persons as ends in themselves, and bestowing priority to individual rights and autonomy. For instance, in What is Sacred, Ronald Dworkin presents an analysis of what he believes is the deeply held conviction in the “sanctity of life”.1 Rather than examining his particular claims regarding how various sorts of life and death practices should be considered, I wish to examine his characterization of life as sacred and the implications that follow. Creating a general conception of the sacred, as presented in the previous chapter, Dworkin argues that “much of what we think about knowledge, experience, art, and nature, for example, presupposes that in different ways these are valuable in themselves and not just for their utility or for the pleasure or satisfaction they bring us. The idea of intrinsic value is commonplace, and it has a

central place in our shared scheme of values and opinions.\textsuperscript{2} He tries to demonstrate the ubiquity of the attribution of intrinsic value by discussing our attitude towards great paintings. We are said to appreciate the inherent quality of art, and not simply a painting’s contribution towards our education or pleasure. He then argues:

If it is a horrible desecration to destroy a painting, for example, even though a painting is not a person, why should it not be a much greater desecration to destroy something whose intrinsic value may be vastly greater?\textsuperscript{3}

As presented in the previous chapter, Dworkin makes two claims in this passage: 1) that we treat both great art and the life of persons as having intrinsic value; and, 2) that we can compare their values. Again, it seems many of us do have the attitude expressed in (1) and a problem arises because it seems many of us also believe that (2) is possible. If we are committed to certain values as intrinsic ones—that is, valuable in themselves and not in terms of any other values, considerations and reasons—how are we able to compare them? Given that we do make such comparisons both in our personal lives and as a society, does this mean that we are choosing randomly?

To answer these questions, we need to understand precisely what is meant by the characterization of treating something as intrinsically valuable.\textsuperscript{4} Unfortunately, Dworkin fails to offer any specification, and indeed it is difficult to do so since the very nature of the concept, as something that is good in itself and not in terms of

\begin{footnotes}
\item[2] Ibid., p. 158
\item[3] Ibid., p. 161
\item[4] In this paper, the notion of intrinsic value, valuing something as an end itself, is being contrasted with instrumental value, valuing something as a means. That is, what is of concern is the way in which we value things, as opposed to the way in which things have value.
\end{footnotes}
anything else, seems to defy specification. However, other philosophers have attempted to give more content to the concept. Therefore in order to flesh out Dworkin’s view, it is useful to examine one such account. As discussed in the previous chapter, Peter Railton offers an account of what it means to treat something as intrinsically valuable and does so more clearly than most philosophers.\(^5\) For Railton what it means to be committed to an end or value in an intrinsic, as opposed to an instrumental, manner is a “matter of (among other things) whether it furnishes one with reasons for acting that are not mediated by other concerns.”\(^6\)

Railton maintains that being committed to a value in an intrinsic manner means that it gives us a direct, unmediated reason to act. Furthermore, he states that this reason need not always override other reasons for acting, for otherwise we would be able to have only very few such commitments. My interest, though, is in cases of conflict between two or more such commitments, what is furnishing us with reason to act. We could say that one’s particular values are still providing unmediated reasons for acting in these circumstances. However, it would not seem to satisfy Railton’s characterization of someone who is *committed* to particular ends to merely maintain that they provide a reason for acting, even any weak reason, without also adding that these reasons motivate one to act for the sake of those ends. In order to count as something that we hold dear, it is not sufficient that a particular value merely furnishes any unmediated reason for acting. An alternative and plausible way to interpret Railton’s criterion, especially given his phrasing throughout that we *act* for


\(^6\) Ibid., p. 101
the sake of our particular ends, is to say that we are motivated or concerned to act directly for the sake of our particular ends.\textsuperscript{7} I am not suggesting that we read Railton’s claim as being that we must always be motivated directly by a particular end in order for that end to count as an intrinsically valuable. That is too strong a position for any plausible view. Rather the claim should be read as maintaining that when we are acting for the sake of a particular end, our motivation for doing so is an unmediated one. But an example will demonstrate how a sophisticated consequentialist’s motivations on behalf of his particular ends will need to ultimately be mediated through other concerns—those that form the level of good consequences.

Say that a sophisticated consequentialist regards both his best friend and his family as intrinsic ends in Railton’s sense that he acts directly for the sake of each. However, when these ends come into conflict, he chooses to act for his family. On what basis does he choose his family? He might choose his family because, to use Railton’s terminology, his family is an overriding end. If his family is an overriding end, however, then at least implicitly, he is comparing his ends. How else can he judge that his commitment to his family is stronger than his commitment to his friend? If he is comparing these commitments, then it seems his concern for his family and his concern for his friend are mediated ultimately by other, more general ends. Perhaps he loves his family more and because of this feels a greater sense of duty to them. But if he is a good sophisticated consequentialist, then he must be choosing to respond to or honor his greater affections to his family, ultimately, by appeal to the general level of good consequences. For even if this level is partly
constituted by the end of devotion to family or loved ones, it would be implausible to assume that his family and friend constitute good consequences without the theory losing its consequentialist character. As I discuss below, there are also further problems with viewing particular ends as constituting more general ones that are independent of the contents of consequentialism.

A second reason he might choose to act for his family over his friend is because circumstances simply favor his sacrificing the time that could be spent for his friend on helping his family. For example, if his friend is on the other side of the world and his family is living nearby, then it will simply be easier, less costly, to help his family and he might therefore be able to help them a lot more. However, the issue of comparability is even more apparent here. When required to make choices between equally held commitments, the sophisticated consequentialist must adjudicate between them based on concerns that are independent of both. When there is conflict between his more particular ends, none of which is overriding, his concern for acting will even more obviously be provided by none of them. He chooses to act for his family because he can do more to help them, but this motivation to help more where he can is not a direct concern for his commitment to his family. He must resort to considerations indirect to, external to, both his family and his friend to decide between them. While he may frequently be consciously motivated by his family and by his friend, he must ultimately, if he is a good consequentialist, mediate these motivations by appealing to the outcomes produced by his decisions. Examining cases of conflict makes this point clearer, but the point does not apply only to
scenarios in which two or more commitments are in immediate conflict, as I discussed in the previous chapter.

I now wish to add that perhaps one way to think about our commitments is to say that they are often weak intrinsic values. They provide reasons for taking some action, but fail to provide reasons for not taking that action. For instance, say that my commitment to my friend furnishes me with a reason for comforting her when she is depressed—call it action X. When X conflicts with Y, the action I would undertake for reasons from my commitment to my work, after some point of engaging in X, I choose to do Y. If I choose to no longer do X because of my pressing need for Y, because I am on the brink of getting fired, my commitment to my friend is failing to provide the reason for not doing X. Some other commitment supplies my reason for not doing X. On the other hand, if my reason for no longer comforting my friend, doing X, is because it seems to be rendering her dependent on me to the point that she is becoming more vulnerable, then my commitment to my friend does supply my reason for no longer doing X. We could describe this as a strong intrinsic value, but only at that time.

It is implausible to think that my commitment to my friend or any of my other particular values would always be a strong intrinsic value. Each may indeed constitute a strong value in some local sense or in some domain of my life. For instance, in the area of my day-to-day projects, my research might strongly override all other undertakings. But even my most worthwhile projects come into conflict with other values, such as when my research conflicts with my health or my duties to my family. We must often weigh and trade against one another even the things we
think of as having intrinsic value, and do so by mediating them through other, more general, values. When we are mediating them through other values then, as I demonstrated in the previous chapter, it is not true that we are making decisions on behalf of our particular values for reasons that are internal to those values. To put this point in terms of the contrast between weak and strong intrinsic values, we can say that strong intrinsic values are only valuable in themselves, but this would not be an accurate description of weak ones.

The end of maximizing good consequences can now be viewed as an intrinsic value in the strong sense, but it does not commit us to disregarding our other serious values. Rather, appealing to the strong end of maximizing good consequences is a means for reflecting on and balancing disparate, strongly-held values. More to the point, serious commitments can provide reasons for acting—if only in the weak sense as described above. Each of our serious particular commitments cannot plausibly always be an intrinsic value in the strong sense even for non-consequentialists. Our commitments, if we have more than one, will sometimes give reasons for acting not mediated by other concerns, but even a consequentialist can have intrinsic values in this sense.

When two serious commitments conflict, having to choose between them may be difficult to say the least, but comparing them by resorting to external reasons does not render one cold, indifferent, or even alienated. Rather, it is an inescapable burden of the condition of scarcity. The main difference between most consequentialist theories and most non-consequentialist theories in terms of the way in which commitments are viewed is that in the case of consequentialism it is decided prior to
serious clashes of commitments what would be a general reason for choosing one commitment over another.\textsuperscript{8} I doubt that anyone would want to fix a charge as serious as alienation on \textit{when} in the decision-making process a standard for choice is invoked. In fact, as suggested in the previous chapter, deciding on a general value for adjudicating particular ones in advance forces one to consider how each particular value might be related to other values and thus encourages broad reflection one’s values even in the absence of obvious conflict between values.

To revisit Dworkin’s claims about the intrinsic value of life in light of what I have argued above, it seems that we in fact treat persons as intrinsic values in the weak sense, given that we often choose not only between persons and other values that we feel are very important, but also between persons and other persons, and must do so by appeal to external reasons. Even Dworkin implicitly provides a reason for the intrinsic valuation of art and human life, a reason which is itself a value. The reason, he argues, that we value art and human life is because of the assumed creative and historical processes that give rise to such objects. Thus even aside from their contribution towards utility and pleasure, it is not the case that the value of human life lies only in itself. We can ask why human life is so valuable, and one reason that is provided is because it is the product of either natural or supernatural creative forces of evolution.\textsuperscript{9} Rather than breeding alienation and coldness, an answer such as this seems to connect us to the world and the individuals in the world.

\textsuperscript{8} But note that this does not mean that there is only one measure of what constitutes good consequences.

\textsuperscript{9} Dworkin, pp. 162-164
Perhaps one reason we feel very strongly that certain objects, including life, are valuable only in themselves is because the connection between them and the external reason for their value is not always obvious and direct. Another reason for the attribution of strong intrinsic value to a certain object might be because the way in which it realizes a more general consideration is significant, or because the consideration that it realizes is a particularly important one. However, it is important for the sake of making decisions to recognize that if we have more than one thing of which we speak of in terms of intrinsic value, we will need to look to other considerations to choose between them when they clash. Contrary to the view that many economists hold, though, this does not imply that it must always be the same considerations. Balancing our values does not require that there is a single ranking, or indeed any fixed ranking at all, of our values. How we judge which values are more salient than others can vary with the circumstances of a situation as the particular conditions and circumstances will warrant that certain considerations be more prominent. Nor is it necessary, again contrary to many economists, to reduce all of our values to a single value—even many consequentialists recognize this, as they frequently describe the state of good consequences as consisting of many different values. But most importantly, weighing and balancing our ends by mediating them through other general considerations does not imply the diminished significance of the things we take to be sacred.

In concluding this section and for the purpose of bringing the discussion to bear explicitly on bioethics, I would like to stress the following points: (1) as a society we do not in fact view life as sacred, if believing life is sacred entails that it
offers us reasons for acting in the strong sense; (2) we simply will not be able to hold each life as sacred, given the condition of scarcity; and, (3) holding any life to be sacred, given the condition of scarcity, will hinder us in considering other values and thus prevent us from being able to fully consider all of the options available for a policy or action, including those that are not direct alternatives. The belief that something is intrinsically valuable in the strong sense and is to be protected at all costs can create a different kind of alienation—distancing us from other important values. This last point will be the emphasis of the next section.

4.3 Comparing Values in Bioethics

Railton’s argument that consequentialists can be alienated from rich, unique, and beautiful aspects of life sounds all too familiar to economists and other who employ cost-benefit analysis. Cost-benefit analysis (CBA) has traditionally been conceived of as the attempt to ensure that the benefits of any action or policy outweigh the costs of that action or policy. Many ethicists greet the option of employing cost-benefit analysis (CBA) with repugnance. How can we compare our most precious values to costs? How are we able to place a monetary value on the ephemeral wonders of the environment, on the splendor and richness of art, and especially on the sacredness of life? How can we price that which is priceless? This kind of questioning often leads to the conclusion that CBA and those who employ it are indifferent, and worse, heartless—alienated from what is most valuable in life. But I will demonstrate that there are different and more important reasons for
employing CBA—in its broadest formulation, CBA involves comparing values against other values and doing so by appealing to still other, more general reasons. Thus, CBA can be viewed as a form of consequentialism, without necessarily maintaining that there is only a single strong intrinsic value and, furthermore, that this value is good consequences such as welfare or happiness. Since broad CBA calls for comparing and balancing values with other values, it is susceptible to the charge that it fails to treat our values intrinsically and thus causes alienation. In this section, I will continue my argument against the claim of alienation in the context of CBA.

To motivate this discussion, it will be useful to first demonstrate a seemingly common misperception involved with the use of CBA, one that does not extend to consequentialist frameworks in general. Some ethicists hold the vastly mistaken belief that when CBA is employed, we are reducing values to “economic values”. In Which Way Down the Slippery Slope, Ruth Macklin epitomizes this misunderstanding when she discusses the possibilities of sliding down a morally problematic slope from making decisions on euthanasia in our society towards the Nazi euthanasia program.\(^\text{10}\) She argues that contrary to others who believe the most dangerous practice is withdrawing treatment from patients who lack the capacity to decide, the greatest danger “lies in the use of an economic rationale”. “When the justification offered for terminating treatment is that it is not ‘costworthy’, or that it is consuming a disproportionate amount of society’s or an institution’s resources, the slide down one of the slopes to the Nazi program has already started”.\(^\text{11}\)

\(^\text{11}\) Ibid., p. 130
But discussing the costs of any policy or action, and employing CBA more generally, is *not* about monetary value. It is not inherently about anything. It is simply a means for comparing our values and the costs of achieving those values. Furthermore, even when CBA is presented in the terms of economic value, it still is not inherently in terms of anything in particular. Monetary measures are just that—measures. When we discuss the monetary cost of terminating life-supporting treatment to a patient for whom doctors will not be able to restore functioning, it is not necessarily the case that the money will otherwise be used to line the pockets of hospital CEOs. Discussing the monetary costs of withdrawing treatment can be shorthand for saying “if we continue to sustain this patient, we will not be able to spend that money on better patient care for other elderly patients, longer post-delivery hospitalization for new mothers, or free immunizations for area immigrants” since there are limited resources. CBA is a means for comparing *uses* of monetary resources against other uses of monetary resources, not for comparing important values against monetary resources.

David Schmidtz makes similar points.\(^\text{12}\) Against the background of using CBA for environmental concerns, Schmidtz argues that CBA is a method for taking our values seriously. For Schmidtz, the strongest argument in favor of employing CBA is that it is a means for introducing accountability. Part of what it means to do proper CBA is to account for all of the direct costs involved with any option,

---

including the costs that are imposed on others, or the external costs.\textsuperscript{13} Often the external costs of an option will be so great, that it will not pass the cost-benefit test, and this implies there is a means for holding people accountable for their decisions involving important values. Many of the points Schmidtz makes in his discussion of CBA are very good, and so I reiterate and expand upon them here. But I also argue that there are important divergences between his view and mine, noting when he fails to go far enough in his defense of CBA, and conversely when he goes too far.

To begin with, both Schmidtz and I maintain that there are two main reasons for employing CBA, although I will argue below that my conception of these reasons allows us to employ CBA to reflect more broadly on our values. These two reasons are the following:

(1) To examine which means of protecting our values will be most cost-effective

and

(2) To explore what important values might be forsaken when favoring others.

The point of the first reason is that it is of course better if we are choosing efficient means for achieving whatever we identify to be our goal, purpose or end. Even if a particular value, such as life, is taken for granted as our end, we still need to consider different means of protecting it. Some means will be more cost-effective than others and so will enable us to either protect more life or instead divert the spare resources to the achievement of other values. This latter possibility raises an important consideration often misunderstood by philosophers, perhaps even by

\textsuperscript{13} To give an example relevant to bioethics, when a woman is considering the costs associated with terminating a pregnancy, the costs to the father might be considered an external cost, both in the case that she decides to continue the pregnancy and in the case that she decides to terminate the pregnancy.
Schmidtz. This point is that even though cost-benefit analysis employs a maximizing rationality, this does not imply that our aim must be to maximize some value. Schmidtz argues that even if CBA considerations call for us to maximize a value, this would not imply that we ought to do so. Sometimes, he argues, such as when confronted with the possibility of being able to save five patients by providing them the organs taken from one patient, we should merely respect a value rather than promote or maximize it.\footnote{Ibid., p. 482} Respecting, as opposed to maximizing, life would imply that we should not sacrifice one patient to save five. But Schmidtz mistakenly believes that it would be despite CBA, and because of other moral considerations, that we should resist maximization. However, CBA, outside of its use in the context of any particular moral theory, only implies that our aim is to minimize some cost, not to maximize some value.\footnote{Allen Buchanan has suggested to me that it might be viewed as having an even more minimal function: to simply allow us to know what the costs of any action are.}

Importantly, exactly what we wish to achieve by minimizing costs is not entailed by the use of this kind of maximizing rationality. It is entirely up to us and will depend on the values we take to be most important. By minimizing costs, the saved resources can be applied either towards securing more of the same value, or of fulfilling some other value. For example, if our end is to secure the procurement of organs and tissues, a default policy that treats all people as potential donors upon death, unless people actively opt-out, might be a considerably lower cost means than using animal organs, which is often ineffective and raises a number of other moral considerations involving the status of animals. Also, preventative care may be a far
more cost-effective way of prolonging life than life-sustaining treatment. In both cases, with the resources that are saved we can either spend them on achieving more of the same end, or divert them towards other values.

The other reason for employing CBA, the second reason stated above, is that it forces us to stop and examine the costs of not choosing other values and ends when we decide to favor certain ones. In other words, it is a method of comparing not only ends with means, but also of comparing ends with other ends. Schmidtz also makes this point, and in making it, he implicitly invokes the notion of opportunity costs, but he does not go far enough in his position. He argues that our values will sometimes come into conflict and we will sometimes be forced to compare them and choose between them. There are a few things to say in response to Schmidtz’s characterization of the problem. First, because we live in a world with limited resources, and indeed simply because time is limited, whenever we choose some policy other opportunities are sacrificed. If we take opportunity costs seriously, whenever we decide to apply resources to some end, not only do we need to consider the obvious alternatives to those resources, those alternatives that are in direct competition, but we should also consider the less obvious alternatives. We should resist the quixotic temptation for believing that those problems that are immediately before our eyes are the only problems that merit our moral attention. For instance in taking the action of prolonging life sustaining treatment for some patient, the opportunity costs are not only the costs of not prolonging life sustaining treatment for other patients. Some of the other forgone opportunities that are not as obvious might be better patient care for other elderly patients, longer post-delivery hospitalization
for new mothers, or free life-saving immunizations for area immigrants. These are all important considerations and so illustrate the need to take (all) forgone values seriously when making decisions about actions or policy. Importantly, though, there is nothing in the concept of opportunity cost that implies that competing alternatives need to be compared based on questions of efficiency. Another policy might be considered on grounds of what is more fair or equitable, or any number of other values.

A second thing to notice about how Schmidtz characterizes the problem is that he seems to believe that what we are forced to compare are the means or methods employed to bring about intrinsic values. They are, in other words, relatively instrumental values that Schmidtz sees as being the objects of comparison and decision-making. For instance, his approach appropriately calls on us to think about the costs and benefits for pollution when considering a particular method of alleviating pollution, such as recycling. But it does not challenge us to compare the value of alleviating pollution with other values. Thus, Schmidtz focuses on employing CBA as a method of weighing the direct costs and benefits of a given end or value, like alleviating pollution, and his view does not encourage broad reflection and comparison between competing intrinsic values. A potential reasons he places the emphasis on direct costs, as opposed to the true opportunity costs, of any policy might be that he views values as lying outside the framework of cost-benefit analysis. Recall that Schmidtz suggests that respecting life would call for us to refuse sacrificing one patient to save five. But, as I discussed before, Schmidtz mistakenly believes that our refusal to sacrifice any lives would be despite CBA, and because of
other moral considerations. He views our values as being moral, and therefore, alternatives to CBA, implying that CBA is not about or at least not always about moral values. If this is one’s view, however, then an important question to ask is how we are able to justify the implementation of CBA into any area of ethics. Unless CBA is itself a (thoroughly) moral framework, what warrants its use in ethical decision-making? More to the point, if one sees the comparison of our moral ends as causing alienation from our ends or as causing other significant moral problems, then the employment of something like CBA would seem to compete with taking our values seriously, or with respecting life in Schmidtz’s example. But I argued above and in the previous chapter that there is no good reason to view consequentialist frameworks in this way.

We are often forced to compare our most precious values. This is an unfortunate reality of the scarcity of means and time. Since there are limitations on the resources that we possess to employ towards the realization of our valued ends, some kind of decision-making approach is often needed. Far from reducing the worth of our ends, though, CBA is the enterprise of taking our ends seriously, when there is more than one intrinsic value. It is a means of ensuring that our ends are realized at the lowest costs and that the alternative values are considered, and both purposes speak to the fact that there are many ends, and even many quantities of the same end, that we deem important.
4.4 Beyond Alienation—Other Costs and Benefits of using CBA

A significant advantage of employing CBA as a framework for comparing and balancing values is that it does not require the use of any particular moral theory. Its application can be integrated into any substantive approach. On the other hand, other recent attempts, especially in the field of bioethics, at providing a framework for comparing our values may not so easily be reconciled with all moral theories. Most notably, Tom Beauchamp and James Childress have advocated the use of general guidelines and standards, in the form of principles, to address pressing questions in biomedical ethics.\(^{16}\) Their view of principles is the following:

> Principles are always binding unless they conflict with other obligations. When a conflict of norms occurs, some balance, harmony, or form of equilibrium between two or more norms must be found; or, alternatively, one norm overrides the other.\(^{17}\)

When there are conflicts, they propose a model of balancing the principles for similar reasoning employed in the advocacy of CBA in this paper: “A model of balancing keeps options open without ‘flatly prohibiting’ them.”\(^{18}\) But because they employ four principles as the basis of many more specific rules, without the background of any comprehensive ethical theory, there is a greater likelihood that their approach, rather than the kind of broad CBA argued for here, will be rejected by the advocates of particular ethical theories. Namely, because it provides content,

---


\(^{18}\) Ibid., p. 483
rather than merely being a framework for considering *any* different values, their version of Principalism may more easily come into conflict with theories that disagree over the specific principles chosen, or that reject the emphasis placed on *only* these principles or the reasons for which the principles are important.

Despite the fact that cost-benefit analysis can be construed as the call to consider all of our values, many philosophers and others find that the *traditional* use of CBA is morally problematic in many ways. Historically CBA required only a Pareto improvement, which occurs if someone is made better off without making someone else worse off. This appears to be a non-controversial, minimal ethical requirement that escapes the problems of interpersonal comparisons. If our federal government were to receive a windfall sum of money and could devote those resources to cancer research without thereby lessening the amount of resources directed towards AIDS research, this would seem to be very appealing. However, there are a number of problems with the notion of Pareto improvement, and while some of them are too involved to discuss here, the most significant problem can be summed up briefly: Pareto improvements are rare. Economic changes usually involve winners and losers.

To address this very serious shortcoming to Pareto improvements, economists developed the idea of a *potential* Pareto improvement, which is when the winners can compensate the losers and still be better off than before. In theory, one asks the “winners” how much they would be willing to pay to have a policy and the “losers” how much they would require for compensation in order to accept the policy, with

---

how much one is willing to pay depending upon how strong one’s preference is. In practice, economists use various techniques to infer people’s preferences from their economic choices. The charge that CBA ignores issues of justice and fairness mainly enters with the use of potential Pareto Improvements and the corresponding use of “willingness-to-pay”. To start with, an analysis that relies on “willingness-to-pay” (or “required compensation to be paid”) might be too centered on the preferences of rational humans and would likely exclude animals, infants, the cognitively impaired, and the environment. Second, “willingness-to-pay” often depends on how much one can pay, and thus how much money one has, and on one’s idea of what would be an appropriate price to pay. If a particular policy benefits the poor and harms the rich, what the poor would be willing to pay for the new policy might not be enough to compensate the rich, in which case the policy would not pass the Pareto test. Even though it might not have any net-benefit in this sense, we may still be convinced that the poor would benefit a great deal more than the rich would lose from the policy’s implementation.

Third, and perhaps most importantly, potential Pareto improvements ignore questions of fairness: the compensation considered is only hypothetical, and in reality, some will win and others will lose. If over the long run each person wins as often as she loses, then that a particular policy benefits some and harms others might not render it unfair. But there is a bias in CBA against the preferences of the poor and other beings besides rational humans to whom we would want to confer moral status. It is also questionable whether policy or actions should be based on inferences drawn from people’s economic choices because this mechanism would leave out
perhaps our most valuable preferences. Some of our most valued “preferences”, if they can be called so, such as a preference for lives free of suffering, are difficult, if not impossible, to signal when one buys clothes, gas, and groceries. Furthermore and perhaps most importantly, even our most valued preferences can change upon the deeper reflection and criticism that comes from public debate. So CBA should not be a substitute for public deliberation. Traditional cost-benefit analysis leaves individuals’ values secure from challenges.

Although I believe that many of the objections raised against CBA are valid, it is important not to confuse traditional CBA with the kind of broad CBA advocated in this paper. However, there are certainly limitations to this approach as well that arise precisely because it does not provide specific content to the costs and benefits considered, and here is where my position most clearly diverges from Schmidtz’s. Schmidtz argues that CBA is not a sufficient condition for any decision. Demonstrating that the benefits outweigh the costs will not render an option decisively correct because, he argues, other considerations will matter. For Schmidtz, CBA is a necessary part of decision-making. If an option fails to pass the test of CBA, or if the costs outweigh the benefits, he asserts it is clear that “further discussion is not warranted”. 20 This at first appears to be a modest position. If the costs of an option are greater than the benefits, then how can we seriously consider such an option? However, without giving further content to the “costs” and to the “benefits”, even this proposal goes too far.

20 Ibid., p. 481
Other considerations will matter, of course, and therefore he is right to argue that CBA cannot be a sufficient condition. But other considerations will also make it plausible to remove the status of necessary condition from CBA. If the importance or weight of the costs and benefits are not included in the analysis, then CBA should not even be a necessary condition. For example, if we are considering expensive life-extending care for a late-stage terminally ill patient, should we abandon such action if the costs outweigh the benefits? This is not easy to answer and will depend on whether the benefits, or values favored, are more or less important than the costs, or the forsaken values. The important question to ask is: what are the alternative uses of resources to sustaining life-support for this patient? If the alternative is that many more lives can be saved through free immunizations, then the cost of sustaining life-support for the patient is potentially great. In this case, that the costs outweigh the benefits might imply that further discussion is not warranted. But if the alternative use of resources is to re-carpet a floor in the hospital, and this would have the result that a greater number of people would be happier, simply that the costs outweigh the benefit would not seem to imply that further discussion is not warranted. Even if the losers are losing more than the winners are gaining this would not be decisive, because what the losers are losing might not be as important as what the winners are winning. The full accounting of costs and benefits will depend on what we consider to be the costs and benefits and how different reasons should weigh in on determining how important each is.

The difficult task of course is how to consider the importance of various values. But it is not whether we should engage in cost-benefit analysis at all or ask
what the costs are. Above I discussed the flip side of maximizing any value, and that is minimizing any cost entailed in achieving a valued end. Again, attempting to minimize a cost of attaining a certain end in the field of bioethics would not imply that the saved costs would line the pockets of policy-makers or hospital board-members. By minimizing costs, the saved resources can be applied either towards securing more of the same value, or of fulfilling some other value. Minimizing the costs, or more generally, inquiring into the costs of any action or policy is the crux of cost-benefit analysis that should be embraced in the context of any ethical decision-making, including in bioethics. We must be willing to meet the actuality of scarcity candidly and be willing to discuss costs in order to treat all of our values seriously. Holding life, for instance, as sacred in the sense of always treating it as a strong intrinsic value, one that cannot be compared, and as something to be protected at all costs, might prevent us from being able to take other values, and even other lives, seriously. The unwillingness to balance our values with other values can occasion a different form of alienation from what Railton has illuminated. By faithfully adhering to the intrinsic moral status of certain values, we become estranged from some of the reasons for our dearest values and hence from other important values. We run the risk of living parochial lives that are detached from others in the world for which our neglect of comparing and balancing values has significant consequences.
4.5 Conclusion

The ideals of the intrinsic value of life, respect for individuals, and treating people as ends in themselves are important because they signify a society’s unwillingness to promote the common good at the expense of the individual. They also place emphasis on values, in particular respect for autonomy, that have historically been undervalued by various health-care practices. Furthermore, they offer goals to aspire to. On the other hand, these values must also be vulnerable to evaluation, criticism and comparison with other significant values. Although this may seem obvious to some, few authors in the ethics literature seem to fully appreciate the implications of the condition of scarcity. This is something which economists have taken seriously. Given that resources and time are both limited, we are often, indeed, always according to some views, forced to choose between competing values.

However, related to the problem that CBA might be used as a test for any decision without full consideration of the significance of the costs and benefits, it seems that philosophers and others are right in at least one aspect of their objection to the use of CBA, even of the broad kind advocated here: that those who employ it seem too easily to want to employ it, and something may be lost in the facility of its employment. In particular, it seems there is some real import in using the language of intrinsic value, pricelessness, and incomparability for those things that have especially high value to us, such as life. Not only might it thwart becoming hardened to the fact of scarcity and its corresponding implication that we must, unfortunately,
make choices between our highest values, but speaking of certain values as if they are priceless might also facilitate the formulation of policies that are cognizant of that high valuation. It is important however to be cautious of dismissing intrinsic value-speech altogether one the one hand, and of failing to realize that the use of such language should not exclude the consideration of other valued ends, on the other hand. Otherwise we run the risk of a different kind of alienation—becoming unaware of other important considerations that we might be trading off.
Part III  Rational Choice and Individualism

5  Liberal Respect for Identity? Only for Particular Ones

5.1 Introduction

In recent years, liberalism has been charged with failing to appropriately value identity-groups based on social traits such as race, culture and religion. This charge about the failure to recognize social identity can also be applied to rational choice theory. As discussed in my Introduction, while rational choice theory only explicitly endorses methodological individualism, a large area of rational choice theory at least implicitly is committed to the stronger thesis of normative individualism, or what Pettit calls ‘personalism’, that is prominent in the liberal tradition.\(^1\) Normative individualism, as the name implies, is a doctrine about what constitutes the good. It maintains that whatever is of value about nations, cultures, businesses, families or any other group, is something that is valuable for the individuals they affect. In short, the thesis denies that there are irreducibly collective goods.\(^2\) Thus, a large part of rational choice theory—that which is committed to the liberal thesis of normative individualism—would deny that there is value in social identity if identity is

---

\(^1\) This is especially true of welfare economics. Welfare economics is the branch of economics that is concerned with evaluating different institutions, actions, and policies, on the basis of normative judgments. The single most important analytical tool in this area has been the condition of pareto efficiency, which maintains that a state of the world, X, is morally better than another state of the world, Y, if in X at least one person is better-off and no one is worse-off than in Y. The pareto condition is an individualist account of the good since it only judges the desirability of these institutions, actions, and policies based on how well-off they leave individual persons. For a fuller explanation of these points, see Brennan (pp. 125-6) on the pareto condition and Pettit (pp. 22-30) on personalism or normative individualism.

\(^2\) Goodin argues that the claim that some goods (for instance, cultures, norms and language) are only possible in a society—a non controversial claim—does not show us that once these goods exist their value is not the value for individual persons but rather emergences from some new property of the group or collective itself. (Goodin, 1990, pp. 71-6)
described as only being a property of a group or collective itself. In this chapter, I argue that if liberalism is to value identity groups, it will be constrained by certain liberal principles. Even though these principles are not necessarily entailed by rational choice theory, I will show how satisfying these principles will only support valuing a form of social identity that is individualist. Thus, even though the explicit starting point of the chapter is liberalism, the result is to countenance a form of social identity that is also fully consistent with rational choice theory.

In what might be the most effective argument from the standpoint of liberalism, communitarian critics argue that social identities are vital to the autonomy and constitution of the persons who are members of identity groups and so the failure to grant recognition and rights to identity-groups amounts to a denial of equal respect for those persons. Any authentically liberal view that attempts to account for this charge would have to be subject to certain constraints. Namely, in recognizing some identity-group the account would have to not merely warn against harming the autonomy of individuals in ways that have traditionally mattered to liberals, but would have to specify a procedure to extend rights and privileges to identity-groups that does not itself undermine the liberal sense of autonomy. With this constraint in place, what would a liberal account of respect for social identity look like, and what exactly would such an account be committed to respecting?

---

3 I am not referring to metaphysical discussions about identity or personal identity. Identity refers to traits that are shared by some people and not others, unlike the universal features of personhood, whatever we hold them to be. Another distinction is between personal identities—those that are not socially salient and defined, such as an odd sense of humor—and social identities—such as one’s race, ethnicity, culture, sexual orientation, religion, or disability. The arguments in this chapter do not hinge on there being a sharp distinction, although it is generally the latter kind that is central to concerns over respect.
In this chapter, I explore an ‘inclusive’ principle of equal respect for persons in which we would respect both universal personhood and social identity. The purpose here is not to argue that we have an obligation to recognize any social identity, but rather to explore what implications would follow from the position that liberals must respect identity as part of the respecting of persons. It will be shown that even under this inclusive formulation of respect, we would need to distinguish an identity belonging to particular persons from the general and collective forms of that identity—for instance it would be Mike’s deafness and not deafness in general and Nina’s culture and not the collective form of the culture that she shares with others that would be the appropriate objects of respect. I will show how the need to preserve the liberal sense of autonomy, together with the fact that people have multiple identities at the same time, supports the distinction between particularized and collective identity.

To determine which acts would be about respecting particularized identity versus other identity forms, I provide a counterfactual test. The result is that any action that passes this test would not seem to benefit a given identity at the expense of other social identities an individual may have and it would not compromise the rights and liberties of individual group-members. The account would thus satisfy the autonomy constraint on an authentically liberal view of respect for identity. A major practical implication of the counterfactual test is that collective rights would not be endorsed. Because social identity is almost always viewed in terms of group-membership, valuing identity seems to be at odds with the principles of liberalism, democratic theory, and rational choice theory where persons are the ultimate source
of value.\textsuperscript{4} I will show, however, that respecting social identity would not necessarily ascribe any special status to groups as such.

The arguments I offer regarding the necessity of drawing a distinction between particularized identity and especially the collective form of the identity do not deny the interdependent relations among persons, that people relate to and genuinely care about others, or, especially, that identities are constructed in a dialogical process in which they depend on others and are influenced by shared institutions. The distinction I make is not between personal identities and social ones but between particularized identities and collective ones. There is a consistently liberal way to recognize social identity once we make this latter distinction. At the same time, however, it is a central theme of this chapter to show that even if communitarians are successful in their arguments concerning the value of social identity to persons and their autonomy, there are important limitations on the implications of these arguments.

\textbf{5.2 Identity and Autonomy}

There is a large literature on the psychology and value of social identity in liberal democratic societies.\textsuperscript{5} The liberal conception of respect for persons—based

\textsuperscript{4} It is true that for Kymlicka in particular, the reason we must respect identity-groups such as cultures is because membership in identity-groups provides a context of choice for the individual—thus, cultures for instance are important ultimately because individual choice and autonomy matter. Members of transitional or waning cultures, he argues, are disadvantaged as compared to the majority cultures with respect to a stable context of choice, and so must be accorded special rights. However, Kymlicka’s view would also advocate special rights for the groups as such.

on the broadly Kantian and Rawlsian idea that persons are autonomous, self-governing agents—has come under strong criticism by communitarians who argue that the emphasis on a universal, rational human nature conflicts with our actual psychology.\(^6\)

Writers such as Michael Sandel and Charles Taylor argue that most of us view ourselves and our choices, whether or not we want to, in terms of our (non-universal) social identities such as race, ethnicity, gender, disability and culture, and they argue that these identities are fundamental to who we are as persons because they are fundamental to our capacity to value and choose—to our autonomy—the very thing that liberals value so much. These critics argue that it does not matter that one’s membership in some identity-group is often not chosen and is not reflectively or consciously regarded as significant. In fact, that our social identities are not always available for scrutiny—that they are outside our ability to reject or accept them—is a key part of why they are viewed as being constitutive of who we are.\(^7\) They constitute who we are in large part because they are the fixed frameworks in which

---


we figure out our values and goals, and this is especially true for people whose social identities are non-voluntary, such as those living in traditional societies and those whose identities have a physical component, such as being deaf or having a particular race.

Taking a slightly weaker position, Will Kymlicka describes social identities, especially cultures, as providing background contexts that give our choices and actions meaning. One’s culture is a largely un-chosen and deeply embedded part of who one is and forms the basis from which one determines a conception of the good and, in particular, within certain boundaries of what is considered appropriate, good or valuable. What all of this means, according to the above critics, is that if the concept of a person and of a person’s autonomy essentially includes social identity, then the liberal principle of respect for persons must also include this identity.

While the notion of respect can mean many different things, what it means in this context is to accord certain rights and privileges to identity-groups that have been faced with disadvantage. For instance, it might involve exemptions from public education requirements for traditional societies, financial support for a private school for deaf students, or providing resources for the construction of a mosque in order for

---

8 Kymlicka in *Liberalism, Community and Culture* and *Multicultural Citizenship*

9 There are reasons for believing that certain identities would be more significant to autonomy than others: those that have been met with social obstacles, whether disadvantage, discrimination or both. It seems these identities would be especially salient for one’s outlook and understanding of the social and political world, and therefore especially important to one’s capacity to value. This is because while disadvantaged identities limit some opportunities, they create others. In fact, the idea that identities serve as constraints seems to a large extent to be the reason they are also viewed as providing options. Opportunities, at least meaningful ones, do not appear in a vacuum. They are generated and gain their meaning in relation to certain restrictions. In this way, disadvantaged identities would seem to provide an even more significant “context of choice” than other identities. (See Kymlicka, *Multicultural Citizenship*, pp. 82-84. Kymlicka uses this term to refer to all sorts of cultures and not just those that have been disadvantaged)
a Muslim community to feel that they are fully accepted in a western society—something that, advocates would argue, would not be seen as establishing any religion but rather recognizing one that is at a disadvantage. Importantly in all such cases, the special benefits that the group receives would be seen as a way to accord equal respect to certain social identities that are at a disadvantage either because they are not part of the majority social identity or because they have been discriminated against or both. Thus, the benefits would not be viewed as in fact being special.\(^{10}\)

Even if critics are correct that liberals must respect social identity in these ways, there are certain constraints internal to the liberal tradition that any liberal account would be subject to. A liberal account could not include those social identities that harm others or, more generally, that are not tolerant of others. If some group, for example, believes that all non-group members are morally inferior, liberals would have reason not to recognize that group. Many authors acknowledge that respect should only be extended to identity-groups that are tolerant of other groups.\(^{11}\) This can be called the ‘tolerance constraint’. Of course understanding precisely what it means to be tolerant of others is a contentious matter in liberal and democratic theory. Putting this difficult issue aside for now, it is not just the harm that may come to other groups that must be guarded against.\(^{12}\)

---

10 Kymlicka, *Multicultural Citizenship*, see especially Chapter 5
11 Gutmann, *Identity*, see especially pp 89-112. Liberals would not be committed to respecting morally repugnant identities, such as a deep hatred for other races, no matter how integral to one’s sense of self. If one’s identity is only partially defined by such an element, however, we may be able to respect other elements of the identity without embracing this aspect.
12 There is also reason to believe, as Gutmann argues, that groups that do not respect the autonomy of individual members will likewise not respect other groups.
More important for the goal at hand, it is a minimum constraint on any liberal account of respect that the autonomy of the members of the identity-group must not be harmed in ways that matter to liberals.\textsuperscript{13} In other words, even if one were to agree with communitarian kinds of arguments concerning the importance of social identity to autonomy, a liberal account of respect for identity would need to ensure that the liberal sense of autonomy would not be undermined. As Kymlicka has pointed out, liberals are committed to protecting the freedom of individual members of an identity-group to question and revise the practices of the group and, more specifically, to protecting their basic rights and liberties.\textsuperscript{14} We can call this the ‘liberal autonomy constraint’. It would not be a liberal account if autonomy could be enhanced in ways that would address the communitarian arguments but would be undermined in ways that are important to the liberal tradition. The latter is a necessary condition for any liberal account to be consistent and is what would set it apart from a communitarian view.

But I want to emphasize that a liberal account must do more than merely caution or assert that respecting social identity should not harm those aspects of autonomy important to the liberal tradition, as other leading liberal accounts have only done.\textsuperscript{15} It must be able to specify a way to recognize social identity that does not

\textsuperscript{13} As Kymlicka has pointed out, “one can ensure tolerance between groups without protecting tolerance of individual dissent within each group”, p. 162
\textsuperscript{14} Kymlicka, see especially pp. 152-158.
\textsuperscript{15} Kymlicka eloquently argues that liberals can endorse ‘external protections’—protecting the resources, institutions, and practices of the group against outsiders—but not ‘internal restrictions’—limitations on the liberty of some group-members imposed by other members. The latter would be a violation of the basic rights and liberties of individuals. However, a problem arises because he does not offer any procedure or guidelines for determining whether a particular practice would fall under the label of external protections or internal restrictions. For instance, is the imposition of marriage on young women (or the demand to exempt children from public education requirements) in a particular
itself undermine the liberal sense of autonomy. More to the point, there must be a built-in safeguard against violating the basic rights and liberties of individual group members. Otherwise, the account would fail to offer any assurances that the autonomy of the individuals who are the object of such identity concerns would be enhanced overall rather than diminished, and thus fail to satisfy the expressed aim of the account.

The overwhelming focus in the debate on social identity has been on membership, whether formally or merely associative, in identity-groups that are marked by a shared trait or characteristic. The question of whether these groups ought to be able to collectively pursue their individual interests or collectively exercise their individual rights has not been an especially contentious matter in liberal discourse. Much more contentious is whether an individual’s rights, status, or obligations as a member of an identity-group should ever trump an individual’s rights, status, or obligations as a person. Because social identity is almost always

---

16 As Appiah notes, these collective rights “tend to have more friends...Most people think that it is just fine that Utah or the city of Cambridge or the Catholic church can exercise rights, through the ballot box or (in the case of churches) through whatever consensual internal mechanisms they agree upon.” p. 72-3
viewed in terms of group-membership, valuing identity seems to be at odds with the principles of liberalism and democratic theory where persons are the ultimate source of value. I will show, however, that respecting social identity in such a way that the autonomy constraint is satisfied would not ascribe any special status to groups as such.

Before demonstrating this, I must first discuss what it would mean to respect social identity as part of the respecting of persons. To sum up the communitarian position, social identity is considered to be significant to who one really is. As such, it is held to be an important source of claims on liberal societies. Bhikhu Parekh puts the claim succinctly when he writes that ‘the liberal is in theory committed to equal respect for persons. Since human beings are culturally embedded, respect for them entails respect for their cultures and ways of life.”

5.3 Including Identity in Respect for Persons

So what would it mean to respect a culture, race, or “way of life” because it is important to who someone is, to her autonomy? As suggested, it would mean including these identities as part of the respecting of persons. Under such an

---

17 It is true that for Kymlicka in particular, the reason we must respect identity-groups such as cultures is because membership in identity-groups provides a context of choice for the individual—thus, cultures for instance are important ultimately because individual choice and autonomy matter. Members of transitional or waning cultures, he argues, are disadvantaged as compared to the majority cultures with respect to a stable context of choice, and so must be accorded special rights. However, Kymlicka’s view would also advocate special rights for the groups as such.

18 This is the view of several prominent political theorists. For instance, writing in regard to culture, Kymlicka argues that the liberal principle of equal respect for persons requires the recognition of the rights of certain cultures since culture is considered “a constitutive part of who the person is” (Multicultural Citizenship p. 175) and Taylor argues likewise that an important condition of respecting someone is recognizing that person as being constituted by her culture. (1992) Multiculturalism and the Politics of Recognition. Princeton: Princeton University Press)
‘Inclusive Principle of Respect for Persons’, or Inclusive Principle for short, individuals who are persons would be respected for their primary personhood, where personhood includes the liberal sense of autonomy that all persons have in common. In addition, in order to fully respect persons, we would also, in the same basic way in which personhood is respected, respect social identity. This is the position that communitarian writers and advocates of minority rights are committed to adopt when they argue that by failing to recognize the identities of certain groups of people, liberals disrespect those people as persons; not only their identities and, especially, not merely in some way. Since the claim is that a social identity is inseparable from who a person is and therefore is to be respected in the same way in which personhood is, instead of considering respect for identities as a separate moral demand, we can view it as a different mode of respect for persons. The two different modes of the Inclusive Principle are:

(1st) Persons must be respected for their personhood
(2nd) Persons must be respected for their social identities

I must immediately state what I am not attempting to do here. First, I am not seeking to provide the conditions for picking out the social identities that would be included. It does seem plausible to include only certain categories of social

---

19 Sometimes when identity-groups, or those that advocate on behalf of them, argue that they are being denied equality or respect, they do not explicitly use the terminology of ‘equal respect as persons’. But it seems clear that their objection is at bottom about being denied equal respect as persons, and again not that they are being disrespected merely in some way. The real concern generally seems to be that members of some identity-group are being deprived of a basic or fundamental respect—that of respect for persons.

20 For instance, it seems that there is a first-person authority concerning what identities are constitutive, but that there must also be some public epistemic constraints, such as that the individual must actually possess the identity as seen third-personally. Questions such as who determines whether an identity is constitutive (and furthermore, meaningfully constitutive) are not explored here.
identities, such as again those that have been met with discrimination or disadvantage and those that do not harm others, but this issue is beyond the scope of the chapter. Second, and more importantly, even once we have such conditions in place for which identities qualify, I am not arguing that the value of social identity is of equal weight with the value of persons for their primary personhood, even if the identities that qualify should all be valued equally with one another.\textsuperscript{21} Since this is a liberal account, it endorses the traditional liberal assumption that for an individual who is a person, her primary moral status is her personhood, universally defined. Social identity is derivative of and dependent on personhood in at least two ways. In the first place, we could not even say that social identity is central to who we are as persons were it not for the personhood features already in place, so to speak. Personhood is constitutively or causally primary—being someone who values and chooses at all is prior to being someone who values and chooses in the particular way that communitarians maintain. More importantly, under the kind of communitarian charge presented, the reason that liberals would be committed to recognizing social identities is, again, because of the persons they comprise.

The primacy of personhood is the basis for the tolerance constraint. Respecting some identity-group cannot violate the 1\textsuperscript{st} mode of respecting persons by harming non-group members or by depriving them of their basic rights and liberties.

\textsuperscript{21} Although there might be important distinctions between respecting someone and valuing someone, I will assume that both notions refer to a basic sense of respect that is involved with the notion of respecting persons—which might be called recognition—and not a more robust or demanding sense of valuing, such as that which might be involved with esteem, admiration, or honor. Furthermore, in some formulations of respect, respect includes the notion of valuing. Respecting any object simply means that the object must be valued in an appropriate way, and not promoted or used. For instance, see Philip Pettit’s discussion in his “Consequentialism and Respect for Persons,” 1989. Ethics 100, pp 116-126.
However, there is a trickier, and for the goals at issue in this chapter, more critical way in which the 1st mode of respecting persons must not be violated; as discussed before, respecting some identity-group cannot deprive the individual members of that group of their basic rights and liberties. In the next section, I will begin to demonstrate how the account here, by maintaining distinctions in identity form, satisfies the liberal autonomy constraint.

5.4 The Locality of Respect

While the Inclusive Principle does not imply that social identity is as valuable as personhood itself, it does suggest that they are valued in the same basic way because both social identity and personhood are important features of persons. This point brings out an important aspect of respect for persons that must now be made explicit: only traits or features of persons would be included. This might seem a trivial point. Of course something like the Grand Canyon would not be included because the principle is about respect for persons. However, in order to understand more about the form of identity that would be included, it is helpful to express why other objects would be excluded. Importantly, it would not be due to the worth or significance of the object to us or to individuals’ lives. The Grand Canyon may have extraordinary value. Instead, the reason that it would not be included in respect for persons is that, simply put, it is not a trait or feature of individuals who are persons.  

22 Not only would something like the Grand Canyon be excluded from the respecting of persons, but it would also seem to be excluded from having the type of value that persons have. This separate claim is not strictly contained in the idea of respect for persons—we could after all hold that both persons and the Grand Canyon have the same kind of worth—but it seems to be related in an important way. In order for the idea that persons are equally worthy of respect to be of significance, it seems only persons
Respect for persons does not directly include respect for anything other than persons, and persons are their traits and features. This does not mean that all traits and features of persons would count for respect, but rather that only these would count.\textsuperscript{23}

This claim applies straightforwardly to social identities that we deem valuable, even highly valuable, but which are not features of persons. Claiming something to be valuable or important to persons is different than claiming something to be an important part of persons.\textsuperscript{24} This is in fact a primary reason that social identity takes on the significance it does: identity is not just important to persons. In popular and scholarly discussions, social identities are seen as being an important part of persons. However, there is a crucial qualification that must be made to the last point: an identity must belong to individuals who are persons in order to be considered a trait or feature of persons and thus to be included in respect for persons.

While it may seem obvious that only identities belonging to persons would be included, there are distinctions in form that identity can take. What I am contrasting is particularized identities from more general types of identities. To say that something ‘belongs to’ or is a ‘part of’ persons is to say that it belongs to or is a part of particular persons. Identities in general or collective identities on the other hand, must be valuable in the way that they are. The force of the principle comes in large part from the singling out of persons as opposed to other objects of value. Even though other objects would not be included in the principle, this does not imply of course that they do not have some other significant value. To take just one example, many non-human animals that do not qualify as persons still have an important value and should not be subject to cruelty and inhumane treatment. Furthermore, this does not imply that the value of persons is incomparable with the value of non-persons or other things. In fact, to say that the Grand Canyon is not of equal value with persons is to compare the its value to the value of persons.

\textsuperscript{23} Dave Estlund gave me a nice example here: personhood respect does not entail fingernail respect.
\textsuperscript{24} These concepts do come apart. Art, family, and a deep passion for traveling can be very valuable to us without being part of anyone’s identity. At the same time, we can have features of our identity that are trivial or even shameful. While we can consistently and perpetually be ashamed of something that is a part of us, it seems the same does not hold for something we value; we would come to no longer value it.
do not belong to anyone in particular. So, for example, if Mike is deaf, Mike’s deafness or Mike’s being deaf would count but deafness in general or deafness as common to members of the deaf community would not. Only particularized identity—Mike’s deafness—is localized enough to persons to count in respect for persons.

I believe that there is intuitive support for deconstructing identity in this way, as I will try to show here and in the next section. However, even if one were tempted to deny these distinctions, I will discuss how maintaining them, and taking actions to benefit identity accordingly, provides a means of recognizing identity that satisfies the liberal autonomy constraint. While it may not be the only conceivable account of respecting identity that would satisfy this constraint, I will show that the account offered here fully satisfies the autonomy constraint, genuinely advances identity-based interests, and does not make, I believe, implausible distinctions in identity. Thus there are several, independent, reasons to endorse these distinctions.

The distinction between a particularized identity, on the one hand, and the collective or general form of the identity, on the other, proceeds from two related claims. The first, and mainly normative, claim is that one’s attitude towards one’s identity is partly constitutive of that identity. The second, and descriptive, claim is that a given identity belonging to an individual will interact with the other identities that she has. Taking the first claim, it seems a liberal account of the importance of identity should want to maintain that someone’s identity will depend not only on the

---

25 There may be important distinctions that must be made between a collective identity and an identity in the abstract, but these distinctions are not necessary for my central purpose here of distinguishing the identities belonging to persons from the less particularistic forms.
ascription of an identity to her but also on her self-conception of that identity. We can recognize that whether one has a certain culture, race or ethnicity is largely beyond one’s control and, because of this involuntariness, will be important to autonomy—this is the communitarian point. However, this does not commit us to the view that how one conceives of that identity is beyond one’s control. Indeed, any liberal autonomy-based account should resist the latter view in order to ensure that the liberal sense of autonomy is preserved. It is important that any liberal account of identity place emphasis on an individual’s own formulation, revision, and valuation of an identity. Maintaining that one’s attitude towards one’s identity is partly constitutive of that identity achieves this emphasis.

Talk of the importance of identity typically takes on the language of how an individual belongs to an ethnicity or culture or race. But there is also a significant sense of belonging, captured in the liberal view of autonomy, which goes in the other direction. There is a sense in which an ethnicity, culture or race belongs to an individual because of that individual’s distinctive conceptualization of the identity. This is supported by the second (descriptive) claim above—the fact that people have multiple identities at the same time. Although someone’s given identity may be considerably defined in a way that also defines the relevant identity-group, it would not seem to comprehensively define her identity. There will be other identities that are important to her conception of a given identity, whether these other identities are shared with other people or highly personal. Different identities might range in their significance to the autonomy and composition of a person, and communitarians may be correct that one’s culture would be especially significant, perhaps much more so.
than one’s gender, for instance. However, the fact of multiple identities means that a given identity belonging to an individual takes on a unique character because it is combined and integrated with that individual’s other identities, even if the identity at issue is a defining or central one of the individual.

Amartya Sen has said that he is, at the same time:

“an Asian, an Indian citizen… an American or British resident, an economist, a dabbler in philosophy, an author… a strong believer in secularism and democracy, a man, a feminist, a hetero-sexual, a defender of gay and lesbian rights, with a nonreligious lifestyle, from a Hindu background, a non-Brahmin, and a nonbeliever in an afterlife (and also, in case the question is asked, a non-believer in a “before-life” as well).”

Importantly, this sort of amalgam of identities seems to apply to all of us. It is not only a feature of modern societies and the mixing of identities they involve. A mother in the traditional village of Kasur, for instance, is, at the same time, a woman, a parent, a mother, a Muslim, a Punjabi speaker, a member of a particular family, of the province of Punjab, a Sunni, a Kasuri, and a Pakistani. Of course, this description only captures (some of) her public identities. Were we to consider her personal beliefs, preferences, habits, and lifestyle, we could say much more. Even if it were true that being a woman is not as significant to who she is as is being a Pakistani, the fact that she is also the former means that her experiences and conception of what it is to be Pakistani will most likely not match those of a man who lives in Pakistan, and the same point can be made concerning the effect of her other identities on her identity as a Pakistani. Furthermore, the assorted identities someone has will vary in their salience depending on the context. In a conflict with India, being a Pakistani

---

might take prominence, but in the matter of “honor killings” on the sub-continent being a woman might take priority over being a Pakistani. So different people with the same identity can prioritize, value, or conceive of that identity differently because of the other identities they have. Drawing the distinctions in identity form in the way that I have fully allows for these differences between individuals’ attitudes towards their identities.

To better understand the distinctions in identity form and why only particularized identities would be included in respect for persons, consider an example. Nina’s South Asian heritage is integral to who she is. She has grown up to think of herself as South Asian and her heritage provides a broad context for choices that she makes throughout her life and has shaped her values and goals. However, it is the fact that it is Nina’s heritage, that makes it integral to her identity as a person, to her autonomy, and not that it is just any South Asian heritage or a South Asian heritage more generally. Notice that we would not consider Aysha’s South Asian heritage to be an important part of Nina. Nina’s attitude towards her heritage will almost certainly be different than Aysha’s attitude towards hers. The fact that each person has several other identities at the same time means that the heritage of each takes on a particularized character because of its interaction with those other identities. Furthermore, each person’s conception of what it means to be South-Asian is likely to be shaped by how relevant and salient being a South-Asian has been in terms of her history and development. Thus, the identity belonging to Nina is distinctive to her self-conception because of how her identity has figured in her history and development and the way in which it interacts with her many other
identities. All of this changes the nature of the identity. It particularizes it and makes it a part of who Nina is as a person in some basic sense.

This is not to deny that a South Asian heritage in general can be very valuable to Nina in a number of ways. The forms of identity range in the role they play in a person’s practical orientation and sense of self. Nina might think it very important that there continue to be people from South Asia and she might value the art, food, music and languages of that region. And mutual identification with others who are also South Asian may be a sufficiently strong form of identity to share certain ends and values, and also to use their shared identity to secure political ends such as obtaining greater resources and ensuring fair and equal treatment. Thus even the more collective form, ‘our South Asian-ness’, is important to Nina. But for the purpose of including social identity as part of persons, the collective form of South Asian culture would not seem to be sufficient. It is not that the importance Nina places on the collective identity is low, but rather that the collective identity is insufficiently localized to constitute who she is.27

The distinctions in identity form seem even stronger if we turn from culturally or socially conceived identities to physical or biologically rooted ones, because the latter are in a sense located in or attached to individuals and so are even more prone to being uniquely viewed and thus even more localized. If Mike is deaf, it would be his deafness and not the deafness of another person that would seem to be significant to who he and to his autonomy. We would not say that the type of deafness that Mike

---

27 More accurately we might say that Nina’s having of a South Asian heritage would be unique or particular to Nina. It is not the identity per se that changes when it is instantiated in an individual, but rather there is a particular way that individual possesses or embodies the identity. I am grateful to Jerry Levinson for clarifying this point to me.
has but present in another person, Tim, is formative of Mike. Nor would deafness in general be identity-forming of Mike—after all, what seems to be crucial in observing that Tim’s deafness would not be a part of Mike is not that the deafness belongs to Tim, but rather that the deafness does not belong to Mike. Mike will have his own experiences of what it means to be deaf, even if it is largely shared with others who are also deaf.

The preceding discussion does not suggest that a person’s social identity must comprise the whole of an individual in order to be particularized and to qualify for being recognized in the principle of respect. On the contrary, as discussed above, people have not one but many identities at any given time; some of them social and some of them highly personal. This is an important psychological fact about persons and their identities that must be accounted for in any account of respect for identity, and especially any liberal one. A social identity takes on the particularized character it does when it belongs to a person in large part because it is combined with, and stands in a particular relation to, other identities of that person. A given identity will only be one feature of a person, possibly a central or significant one, but must be one that is particularized to a person if it is to count here. Social identity that is not particularized to persons would be excluded from the respecting of persons in a similar way that the Grand Canyon would be. For while it might be something we deeply value, it would not constitute a particular person.
5.5 Collective vs. Particularized Identities

Many of the social identities that are central in the debate and politics of respect for identity may seem to complicate the contrast that I have drawn between particularized and collective identity. Culture is the most obvious example. Someone might argue that the collective culture is her identity. Such a person would not need to maintain that the collective culture makes-up the whole of who she is—only that the collective culture is her culture—that part of her whole person that is made up by her culture. So, this objector claims, the only way for her cultural identity to be respected is to respect the collective cultural identity. Rather, the collective identity cannot even be meaningfully distinguished from the particularized identity, at least in terms of applying the principle of respect for persons.

In order to be included in respect for persons, it should be emphasized that the collective identity would indeed need to be equivalent to the particular identities of persons who are members of that identity-group. In other words, it is not sufficient to maintain merely that there are overlaps and features in common between the collective culture and the culture belonging to Nina. This weaker position would be perfectly compatible with the distinction I have drawn. In order to collapse the distinction then, one would need to say that the collective cultural identity is identical to the cultural identity of particular persons. However, there are problems with this equivalence.

28 Again, one would not need to maintain that the collective identity in question is equivalent to the whole identities of particular persons, only that the given collective identity is equivalent to the counterpart identity in those persons.
One way to summarize the normative claim made above regarding the liberal sense of autonomy is to caution, as Anthony Appiah has, that we could be trading one tyranny for another—that of the larger society for that of the group.29 Emphasizing the collective nature of some social identity may liberate the identity-group from the norms of the larger society, but can also unduly subject individual members of the group to the norms of the group, and in particular, to how the group construes their shared identity. Even apart from this distinctively liberal response, though, it seems collective and particularized identities can and do come apart, not only conceptually but also practically, as I will briefly try to show here and will discuss at greater length in the next section.

In popular and political discussions, people often point out that disrespecting someone who is like them, but not them specifically, makes them feel that their own identity is being disrespected. However, while it is common to feel if one disrespects someone who is like me, but not me specifically, that my own identity is implicated, this seems less true in scenarios in which one respects or benefits someone like me but not me specifically. If Aysha’s boss were to grant her special time off for celebrating a South Asian holiday, this would unlikely make Nina, who works at a different firm, feel as if the South Asian heritage belonging to her was being benefited, at least not to the extent that it was for Aysha. Even if Nina were to know

---

29 Appiah eloquently warns “the politics of recognition, if pursued with excessive zeal, can seem to require that one’s skin color, one’s sexual body, should be politically acknowledged in ways that make it hard for those who want to treat their skin and their sexual body as personal dimensions of the self. And personal, here does not mean secret or (per impossible) wholly unscripted or innocent of social meanings; it means, rather, something that is not too tightly scripted, not too resistant to our individual vagaries.” (original emphasis) Appiah, p.110. However, since he does not explicitly consider the distinction I make between a general type of identity and an identity belonging to persons, he worries that any view that supports recognition for one’s identity is susceptible to the charge of social imposition.
that she would also get time off if she were working for Aysha’s boss, it would still be the case that Nina’s South Asian heritage would not currently be benefited in the same way or to the extent that it was for Aysha.

It could be objected that I have only shown that Nina’s whole person is not being benefited to the extent that Aysha’s is, but that I have not shown that Nina’s cultural identity is not receiving the same benefit. In other words, it could be said that what I have demonstrated is that it is the other parts of Nina, other than her culture, that are not being benefited, and this is why there is a difference in the way that Aysha and Nina experience the action of Aysha’s boss. But it seems that if this were true, then a good case could be made that one’s culture is not a very significant part of one’s whole person. For if the other parts of Nina are such that they are capable of diminishing or countervailing the presumed beneficial effect on her cultural identity to the point that Nina as a whole does not benefit, then this serves to undermine the claim that culture is basic to who Nina is as a person. On the other hand, my claim that Nina would not benefit to the extent that Aysha would does not rest on the idea that other parts of Nina would counteract the effects on the part of her that is her culture. My claim is that Nina’s *culture* would not benefit to the extent that Aysha’s would and *this* is why Nina would experience the action of Aysha’s boss differently than Aysha would. This claim does not undermine the point that Nina’s culture is basic to who she is as a person and in fact supports it.

There are of course many examples of actions that would equally express respect for both the particularized and collective forms of the identity at the same time, and I will say more about this in the next section. But the point here is that
there also seem to be examples of actions that would express respect for one person’s identity without doing so, or doing so to a lesser degree, for another person’s identity—and this is what is needed to establish that particularized and collective identities come apart. To take the analogy with personhood features, suppose that Bob’s capacity to value was being respected in some very generic and minimal way. His neighbor tells him that he admires the bare fact that Bob values things. Even without referring to some special or intense way that Bob values, his neighbor’s comments would unlikely demonstrate respect for Carol’s capacity to value in the same way or measure as it does for Bob’s. And it seems social identities are more capable of being unique to the individuals who possess them than are personhood features. Social identities are richer than personhood features and therefore allow greater room for variation, hybridization, and interpretation across individuals. While the features of personhood are common enough to apply universally to all persons, social identities of course are not.

It might seem that I have only illustrated the gap between an identity belonging to one person from the identity belonging to another person, and not between a particularized identity and a collective one. This would be an important objection because the point of saying that one’s culture is one’s identity is not to stress that the identities of different members of an identity-group are equivalent, but to maintain that the collective identity is indistinguishable from the identity of the persons who belong to the group. However, the fact that one member’s identity is different, even if only in small ways, from another member’s identity suggests that there is a collective identity that stands separate from both.
If we say that the collective form of South-Asian culture is Nina’s cultural identity, and Aysha is also a member of this culture, than we would have to say that the collective identity is also Aysha’s identity. Otherwise, we would be maintaining the odd view that while Nina’s identity is the collective one Aysha’s is not. Someone who takes the hard-line view of claiming that a collective identity is indistinguishable from the identity of someone who is a member of that group would not be maintaining that only one member’s identity is the collective one, of course, but that every member’s is the collective one. So the collective identity would be indistinguishable from both members’ identities. If this were the case, however, it would mean that we could not meaningfully distinguish between one member’s identity and another member’s identity. But I have tried to show that we in fact can, largely because of the role of multiple identities. If we can make this distinction, then there is a collective form of the identity that is also separate from both these members’ identities. The fact of multiple identities and the particularized fusions they give rise to does not mean that we would need to respect each instantiation of an identity in some unique fashion—that would obviously be very difficult to do. It does suggest, however, that there is a gap between the identity of one person and that of another, and thus between a collective and particularized one.

None of the preceding discussion denies that persons depend on others for the construction of their identities or that people genuinely care about others. Political theorists such as Sandel and Taylor are correct to argue that people value relationships of mutual identification and support. Furthermore, the discussion has not attempted to refute the claim that many of our identities are inherently social, in
both their causes and constituents, as it is these that are central in concerns over identity. Nina’s South Asian heritage is of course social in nature and dependent on a shared history—what would it even mean otherwise? But it would still only be the social identity belonging to Nina that would be included in respect for persons. The distinction I have drawn is between particularized identities and collective ones, not between personal identities and social ones.

As I said above, I believe there is intuitive support for unpacking identity in the way I that I have. However, even if one were tempted to deny these distinctions, I have tried to show the liberal rationale for maintaining them. In the next section, I will discuss how taking actions on behalf of particularized identity provides a means of benefiting or promoting identity that satisfies the liberal autonomy constraint, yet still genuinely recognizes identity.

5.6 Practical Implications of the Distinction

To restate the central argument, if liberals are committed to respecting social identity as part of what it means to respect persons, the commitment would be limited to recognizing social identities belonging to particular persons and not general or collective ones. This is because it is only particularized identities that would be constitutive of persons and their autonomy. I argued that there is a conceptual and normative gap between particularized identities and, especially, collective ones, but there is also a practical gap between them.

We can begin by asking what kinds of actions would benefit particularized identities as opposed to other forms. It might seem that pragmatically speaking there
would be no difference; whatever action benefits social identities belonging to particular persons does so because it promotes the broader category. However, there is a useful counterfactual test for determining which actions would benefit particularized identities without necessarily benefiting the collective or general form: that of examining whether the action would benefit or promote an identity-based interest if only one individual were affected by it or, more generally, if the action would refer to only one instance of the identity in question. That this kind of query is intelligible also provides us with another reason for making the conceptual distinction between particularized and collective identity. Again, they do come apart in important and meaningful ways.

It should be stated that I am not urging that we are required to take actions that pass this test for respecting social identity, nor that all such actions would be cases related to the concept of respect, as opposed to other moral or political demands. Rather, my concern is in determining which actions would be justified in terms of particularized identities. Proposed actions that pass the counterfactual test would represent a prima facie identity-based reason for action. Whether we ought to take such an action or implement it into policy would depend on a number of conditions, such as practical feasibility, the efficacy of the action in promoting significant identity-based interests, costs, and the need to avoid disrespecting the identities of other groups. Furthermore, even though the test specifies that an action must be able to refer to only one instance of the identity, the purpose of implementing the action would of course be to benefit several people who possess that identity—to benefit an identity-group.
To see how the test would operate, consider expanding the current policy under the American with Disabilities Act of requiring wheelchair accessibility in public buildings to pertain to all private buildings as well. Even if there were only one person living who required a wheelchair, this policy would benefit the identity in question. Although it may require a community of wheelchair users to warrant the associated costs, and it would of course potentially benefit every member of this community, the action would not require such a community in order for it to confer respect to someone who depends on a wheelchair. In this way, we can see that the policy would be justified in terms of the particularized identity. It may also affect the collective identity of course, but even in the absence of others with the relevant type of identity, the action would still promote an identity-based interest. To take another example, a policy that allowed Muslims to take five breaks during the workday, for the purpose of prayer, would benefit a single Muslim person. It would be a meaningful expression of recognition even in the absence of other Muslims. Thus the action would be about respecting a Muslim identity belonging to particular persons.

Certain acts of recognition for identity would only be meaningful in the context of an identity-group. Language, which is an important element of many cultural identities, presents a good example since it depends on a community of speakers. Teaching a minority language in school (in addition to the majority language) would really only be beneficial if there were more than one speaker of the language in existence. In the absence of other speakers, there would be little point to
introducing it to anyone. However, there are some actions that would respect a language, such as posting signs written in a minority language, which would benefit the identity if only one person spoke the language, even though it is an implausible counterfactual. None of these examples of actions taken on behalf of particularized identity would be trivial. They each would promote significant identity-based interests.

I said that the distinction I have drawn is between particularized identities and collective ones, not between personal and social identities. The concerns raised over identity almost always pertain to social identities such as culture, race and ethnicity, and so the Inclusive Principle would pertain to social identities, albeit, not in their collective form. However, since the counterfactual test calls for considering actions that must be able to refer to only a single instance of an identity, it might be objected that I have not shown how genuinely social identities, which necessarily involve others, would be included. In response to this charge, we can note that while it is true that we can only make sense of a social identity in the context of others, this does not mean that we can only make sense of actions about a social identity in the context of others.

For an action to respect a social identity, it must promote interests of a certain kind: those that individuals have as a result of being part of a group or society. Under the counterfactual test, an action must be able to benefit any single individual’s interests, but this would include those interests that arise out of membership in an identity-group. The concern to pray five times a day is something that comes from

---

30 Indeed teaching any language would only be possible in the context of more than one speaker of the language, as the teacher would already have to know the language.
taking part in the Islamic religion. Likewise, one’s interest in having signs written in her minority language is one that is born from being a member of that language community. In this way, both an allowance for breaks for prayer and special signs would benefit social identities.

Furthermore, as I mentioned above, an action would need to do more than just pass the counterfactual test if it has any chance of becoming policy. It would also have to be worth incurring the costs of implementation. In most cases this will mean that the action would potentially benefit a large community. What such an action would do then is to pick out those features of a given identity that members of some identity-group have in common and so would be about the shared features of that identity. It would not promote the individualized features of identity or those that are blended with someone’s other identities.

Finally, even though the justification of these actions would need to be in terms of benefiting particularized identity, this does not mean that the policy would not in practice also promote the collective form. Putting up signs in a minority language would do more than benefit each individual speaker of that language. It would also facilitate coordination and planning of shared activities of the group of speakers of that language, and thereby benefit the collective identity as well. It is not as if the action would have to benefit individuals alone or in private. To the contrary, actions that would be justified in virtue of respecting particularized identity would often enable members of an identity group to come together. Granting Muslims prayer-breaks would benefit a single Muslim, but would also facilitate increased assembly among Muslims that work in close proximity to each other—if they would
be able to take breaks during the workday for prayer, they would have a greater opportunity for meeting with their Muslim co-workers than if no such breaks were allowed, and this would in turn benefit the collective identity of a community of Muslims.

At the same time, though, and this is an important point, collective rights would not be endorsed. There are different categories of collective rights. In one category are rights justified by reference to the interests of more than one individual or to the aggregated interests of each individual to coordinate their activities. An example of this might be preserving a minority language. In a second category are rights that refer to a collective goal, often expressed through institutions of authority, and not merely to the aggregated interests of individuals. Advocates of this second category of collective rights claim that the group has a shared interest in maintaining certain norms even if some individual members of the group do not endorse the practice. Examples include the right of Muslims to enforce the wearing of headscarves by women in their community and, as in the famous Wisconsin vs. Yoder case, a right of elders in a society to exempt their children from government education requirements.

31 Charles Taylor argues, with respect to culture, that in order to protect an individual’s identity we must try to preserve the distinctive tradition of the identity-group, which would imply support for collective rights of certain kinds. (1994) Multiculturalism: Examining the Politics of Recognition, ed. Amy Gutmann. Princeton: Princeton University Press. And Kymlicka argues that we must grant special rights to certain transitional or waning groups and to members of these groups in order to ensure equality among individuals’ contexts for choice. (See Note 16) But as John Tomasi has argued in response to Kymlicka’s view, stable cultures can also sharply constrain one’s context of choice by closing off a wide range of options that are outside of the particular culture. Furthermore, if we eliminate the condition of stability on what counts as a culture and simply maintain an existential interpretation of culture, Tomasi argues that it is not plausible to see anyone as lacking a culture. Even transitional or waning cultures are still cultures in this uninteresting sense, perhaps hybrids of a sort. On this interpretation of culture, it would not be clear how any cultures would be at a disadvantage. (“Kymlicka, Liberalism, and Respect for Cultural Minorities”, 1995, Ethics, 105, no. 3)
Both categories of collective rights would be ruled out by the counterfactual test. This means that even if the case could be made that the right in question would belong to individuals, the test would exclude such things as preserving a minority language, because the latter sort of actions would need to refer to the interests of more than one individual with the relevant identity. The test therefore does not simply restate the normative individualism of both liberalism and rational choice theory, as discussed in the Introduction. While the test does require the condition that the rights and privileges in question belong to individuals, it requires more than this. The view here is also stronger than the liberal presumption against harm; for it is not clear that activities such as preserving a language would cause harm to anyone. Thus the test does not merely reiterate existing liberal intuitions.

The test is however constrained by an important liberal principle. Earlier in the chapter I said that any liberal account of respect for social identity must do more than simply caution or assert that an action aimed at recognizing some identity-group should not weaken the autonomy of the individuals in the group. It must be able to specify a way to recognize social identity that does not itself undermine the liberal sense of autonomy, and in particular basic civil and political rights. Otherwise, the

---

32 See Gutmann, chapter 1. It seems appropriate that the condition discussed here would be stronger than the condition that basic rights and liberties belong to individuals. First, whereas basic rights and liberties are universally applied, special identity rights and privileges would of course only apply to some people, and we would need to have assurance that identity recognition would not harm the autonomy of these people relative to others. More to the point, in order to best ensure that recognition for identity would not harm the liberal sense of autonomy, it seems important that such identity rights and benefits not only belong to individuals but also maximally allow for formulating and revising identity conceptions. Exposure to other customs, communities, and cultures is important to that end. Many rights and privileges that would belong to individuals but which would need to refer to the interests of more than one person, such as special language rights and exemptions from public education requirements, would tend to isolate cultures from one another, and so would not be maximally conducive to promoting the liberal sense of autonomy. The view offered here not only provides a means of safeguarding against the violation of basic rights and liberties, but is also responsive to claims about how cultures are not pure, unchanging things.
account could not be said to adequately show how it would enhance rather than diminish autonomy overall—the expressed aim of such an account. This was the liberal autonomy constraint.

I have demonstrated how the account here, with the distinctions in identity form and correlative counterfactual test, specifies such a way. In particular, the counterfactual test provides a practical procedure for ensuring that the liberal sense of autonomy is not placed at peril. Because the test would exclude collective rights, we would have assurances that an action would benefit particularized identity, even if it also happens to simultaneously benefit the collective one in certain ways. Since an action would have to be able to benefit an identity-based interest of any single individual in the absence of others with the identity, it would not promote collective interests at the expense of the basic liberties and rights of individuals with that identity.

This point is another way of appealing to the fact of multiple identities. Under the test, any action that would benefit a given identity would not infringe on the other social identities individuals may have. A right to prayer-breaks would not compromise a Muslim woman’s identity as a woman. The account here thus more than satisfies the liberal autonomy constraint. It does not trade-off the liberal conception of the nature of autonomy for other conceptions.

5.7 Conclusion

Because identity is so frequently, if not always, cashed out in terms of the identity of groups, liberal political theorists tend to view respect for social identity as
giving status to groups as such, and frequently, as giving priority to groups over individuals. My arguments have shown a way in which liberals can consistently respect social identity. The prayer break example demonstrates that a social identity can be actively respected without undermining the basic rights and liberties of individual Muslims. It is an example of an action that would respect the Muslim identity belonging to particular persons. However, an important theme of this chapter has been to show that respecting social identity as part of the respecting of persons would not support more robust communitarian positions endorsing special collective rights and privileges. There may be other important ways of valuing identity, but for the purpose of including it in the concept of a person and a person’s autonomy, we must distinguish between a collective identity and particular individuals’ identities, and include only the latter.
Bibliography


Fehr, Ernst and Fischbacher, Urs. 2004. Social norms and human cooperation, *TRENDS in Cognitive Sciences* Vol.8 No.4, 185-190.


Biography

Sahar Akhtar was born in London, England to Hanif and Zahida Akhtar. She received her bachelors of arts from George Mason University, in economics. After working for several years in the federal government, she went on to complete a doctorate in economics, also from George Mason, in 2001. During 2007-2008, she has been a Post-doctoral Fellow in the Political Theory Project at Brown University while she finalized the present dissertation. In the fall of 2008, she will begin a position as assistant professor in the department of philosophy at the University of Virginia.