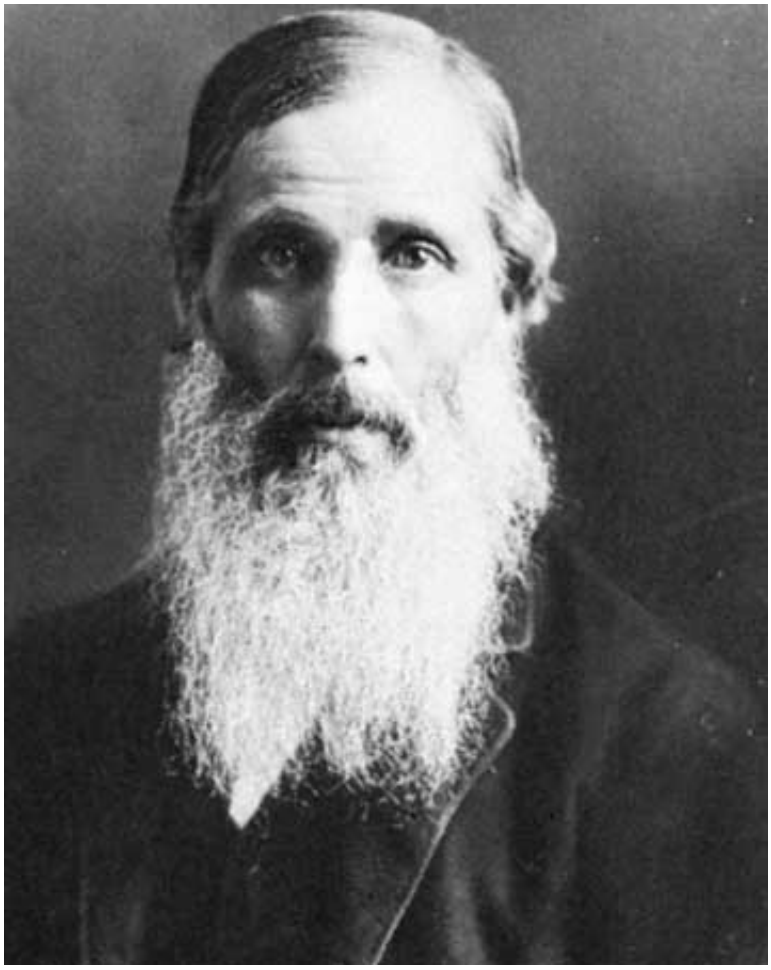
Etica & Politica / Ethics & Politics

X, 2008, 2



Università di Trieste
Dipartimento di Filosofia
www.units.it/etica

Edizioni Università di Trieste

I S S N 1 8 2 5 - 5 1 6 7

Etica & Politica / Ethics & Politics

X, 2008, 2

Etica & Politica / Ethics & Politics
X, 2008, 2

MONOGRAPHICA:
THEMES FROM SIDGWICK

SYMPOSIUM:
WALTER BLOCK, *Labor Economics from a Free Market Perspective*

VARIA

MONOGRAPHICA:
THEMES FROM SIDGWICK

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 9-18

Guest Editors's Preface

Gianfranco Pellegrino
Luiss "Guido Carli"
yorick612@gmail.com

Massimo Renzo
University of Stirling
massimo.renzo@stir.ac.uk

In the following pages the reader will find the first part of a collection of essays devoted to themes from the thought of Henry Sidgwick (1838-1900), mainly focused on his masterpiece, *The Methods of Ethics* (1874).[1] The work of Henry Sidgwick has had certainly a peculiar fate in the philosophical debate of the twentieth century. As lamented by Bart Schultz in the *Foreword* to his classic collection *Essays on Henry Sidgwick*, published in the early nineties, the attention paid to Sidgwick's work is not comparable to the attention received by the great British thinkers of the past. We still do not have critical editions of his work, nor do we have many volumes dedicated to him (there are relatively few indeed if compared to the studies available on Hobbes, Hume or Mill). Finally, at least at the time when Schultz was writing, Sidgwick's books, with the exception of the *Methods of Ethics* and the *Outlines of the History of Ethics for English Readers*, were unobtainable.[2]

It is interesting to note however that in spite of the scant attention received in academia,[3] Sidgwick greatly influenced some of the most important moral and political philosophers of the twentieth century. Philosophers like George Edward Moore, John Rawls and Derek Parfit all acknowledged their debt to him, so that it would not be an exaggeration to claim that Sidgwick played a formative role in setting the agenda and the methodology of our current discussions on metaethics and normative ethics.

---

[1] A second group of contributions will follow in the next issue of *Etica & Politica/Ethics & Politics*.

[2] Schultz 1992.

[3] One of the reasons usually produced to explain the scant attention received by Sidgwick is his writing style, which most people seem to find pedantic and rather dull (see Broad 1930, pp. 143-144; Selby-Bigge 1890, p. 93). For a different opinion see Blanshard 1984, p. 21; Rashdall 1885, p. 200.

This view seems to be confirmed by two other prominent figures in the contemporary debate in moral and political philosophy, David Gauthier and Stephen Toulmin, both of whom argued that Sidgwick, rather than Moore, can be considered the real father of contemporary moral philosophy, since it is in *The Methods of Ethics* that the distinction between normative questions and questions about the meaning and the nature of judgments (which marks the beginning of contemporary metaethics) was explicitly defended for the first time.[4]

This however is not the only reason why Sidgwick can be said to have created "the prototype of the modern treatment of moral philosophy;"[5] or, in Rawls' famous words, "the first truly academic work in moral theory, modern in both method and spirit".[6] Sidgwick is arguably the first philosopher who treats ethics as an autonomous area of investigation, not depending for its conclusions on the acceptance of a particular metaphysical system. In the *Methods of Ethics* he starts instead with the ordinary beliefs of individuals belonging to a specific place and time, and then proceeds by way of a reflective dialogue between these beliefs and some of the most important ethical principles advanced in the history of moral philosophy. The idea underlying this approach is that moral principles can only be founded in the reflective worldviews of the agents who have to recognize and endorse them.[7]

Nussbaum correctly traces back this approach to Aristotle,[8] which is certainly a plausible interpretation, since Sidgwick himself presents his work as an attempt to "imitate" Aristotle's examination of "the Common Sense Morality of Greece, reduced to consistency by careful comparison: given not as something external to him but as what 'we' – he and others – think, ascertained by reflection".[9] However it is only with Sidgwick that it is clearly stated for the first time (at least in the modern era) that the only way to reach an adequate justification in ethics is by a systematic comparison between the different conceptions of morality and the different

---

[4] See Gauthier 1970, p. 7; Toulmin 1986, pp. VII-XX, Hurka 2003. For a different view see Cremaschi 2006.

[5] Schneewind 1977, p. 1.

[6] Rawls 1980, p. 341.

[7] It is controversial whether this approach can be said to anticipate Rawls' method of "reflective equilibrium". For a criticism of this view see Singer 1974.

[8] Nussbaum 1986, pp. 10 and 424, footnote 16.

[9] Sidgwick 1981, p. xxi.

methods that these conceptions presuppose. Hence Sidgwick's attempt to reduce all possible moral theories to three fundamental models: egoism, intuitionism and utilitarianism.

It should be noticed that in Sidgwick this idea is closely connected to another idea; namely the belief that all moral problems can be reduced to fundamental moral questions. Once they are so reduced, according to Sidgwick, moral theories will provide an answer to these problems. This view has been thoroughly criticized by the so-called anti-theorists,[10] but is still widely shared by most moral philosophers working in the Anglo-American tradition.

These are all reasons that bolster Schneewind's conclusion that "Sidgwick gave the problems of ethics the form in which they have dominated British and American moral philosophy since his time"[11], which is in turn echoed by Eugenio Lecaldano's observation "that in the same way in which we can look at Adam Smith – with many simplifications – as the founder of scientific economics, we can look at Sidgwick's work as the first attempt to provide a completely rational and scientific study of ethical conduct".[12]

To this we should add that *The Methods of Ethics* offers a series of insightful theses about the nature of moral judgments and moral concepts,[13] the concept of justice,[14] the critique of moral naturalism and the analysis of hedonism (to name but a few). These theses will constitute a constant point of reference for the contemporary debate. The same is true for Sidgwick's particular formulation of utilitarianism, which is widely acknowledged as the clearest and most sophisticated version of the classical doctrine,[15] and is still one of the most influential in the current debate.[16]

---

[10] See for example Williams 1985; Hampshire 1983 and 1989; Baier 1985; Taylor 1985; Larmore 1987.

[11] Schneewind 1977, p. 422. Some interesting remarks on the merits and defects of Sidgwick's approach to ethics can be found in Rawls 1980, pp. 314-3; but see also Rawls 1981 and 1975. A study of the affinities and the differences between Sidgwick's and Rawls' approach is yet to be produced. For some interesting ideas about this comparison see Barry 1973, pp. 4-9; M.G. Singer 1976; Schultz 1992, pp. 7, 39 and 49-51.

[12] Lecaldano 1996, p. 498.

[13] One of the aspects of Sidgwick's thought which has received more attention is his moral epistemology, which combines a particular form of intuitionism with a sophisticated analysis of common sense morality. See Schneewind 1963; P. Singer 1974; Sverdlik 1985; Brink 1994; Daurio 1997; Pellegrino 2000; Crisp 2002.

[14] According to Herbert Hart, *The Methods of Ethics* (together with Perelman's *De la Justice*) contains "the best modern elucidations of the idea of justice"; see Hart 1994, p. 299.

[15] See Rawls 1981.

[16] On the influence of Sidgwick's utilitarianism on the theories of Richard M. Hare, David Brink, Philip Pettit and Peter Railton see Renzo 2008.

In light of these considerations it should come as no surprise that Sidgwick's work has received more and more attention over the last twenty years. In 1996 a complete edition of his works, including two volumes of essays and reviews not previously collected, have been published by Thoemmes.[17] In 1998 Sissela Bok published a new edition of *Practical Ethics*, drawing attention to the importance of Sidgwick's contribution to this area of ethics.[18] Bok's volume was followed a couple of years later by another collection of Sidgwick's essays, edited by Marcus G. Singer, which highlights the importance of Sidgwick's contribution not only to ethical questions, but also to value theory in general, to moral psychology and to philosophical method.[19] In 2000, for the centenary of Sidgwick's death, *Utilitas* published a special issue on his work,[20] while the British Academy organized a conference whose proceedings were published the following year in a volume edited by Ross Harrison.[21] Finally, in 2006 Bart Schultz published a long-awaited biography which offers an extremely detailed portrait of Sidgwick's life and of his intellectual development, as well as of his political views.[22]

Certainly this is not enough to give Sidgwick a position comparable to that of Hobbes, Hume or Mill in the Olympus of British moral philosophers. Yet the situation is clearly very different from the one described by Schultz in his *Foreword*, more than 15 years ago. Our intention in this issue is to contribute to this renaissance of Sidgwick studies by putting together a collection of articles that explores some of the most important aspects of his thought. The aim is to go beyond the mere rediscovery of a neglected author and to contribute to that mature stage of Sidgwickian scholarship, which will hopefully keep flourishing in the next decades.

Mature scholarship has among its marks a focus on puzzling aspects, rather than a concern with completeness, so we left our authors free to concentrate on those aspects of Sidgwick's thought which most interested them, without any constraint or theme assigned. In selecting the contributors to this collection however we have been guided by three main concerns. First, we wanted the collection to further our understanding of

[17] Sidgwick 1996.
[18] Sidgwick 1998.
[19] Singer 2000.
[20] AA VV 2000.
[21] Harrison 2001.
[22] Schultz 2004.

Sidgwick's ethical thought (see the contributions of Robert Shaver and Anthony Skelton, two well-known Sidgwickian scholars). Second, we wanted to investigate the relationships between his thought and the philosophy of other key figures in the history of philosophy (see the pieces by Sergio Cremaschi, Massimo Reichlin and Alessio Vaccari). Finally, we wanted to show the relevance of Sidgwick's ideas for some of the most important current debates in moral philosophy (see the pieces by Tim Mulgan and Francesco Orsi). Thus this collection aims not only to be a valuable source for those interested in Sidgwick's scholarship, but also to offer a picture of the themes in Sidgwick's philosophy that both contemporary philosophers and historians of philosophy find interesting and worth engaging with.

Not surprisingly Sidgwick's dualism of practical reason confirms its role as one of the themes to which philosophers pay most attention. Francesco Orsi provides a critical survey of the many different readings of the dualism and argues in favour of a specific interpretation according to which Sidgwick's puzzle is not only epistemic or logic, but also practical. Orsi offers an account of the dualism in which egoism and utilitarianism are logically compatible while remaining conflicting principles in terms of "all things considered" reasons. Tim Mulgan focuses on what Sidgwick considered as a possible solution to the dualism (though one he was skeptical about), namely postulating a divine moral order. Mulgan argues that, contrary to what Sidgwick thought, a non-dualistic morality does not require either absolute freedom of the will or believing in eternal survival. Accordingly, morality is less demanding than religion, and no religious premises are needed to overcome the dualism. Finally, Alessio Vaccari describes how the origins of the problem can be found in the dualist ethical theory advocated by Joseph Butler. After comparing Butler's treatment of prudence and morality to Sidgwick's treatment of egoism and morality, Vaccari considers whether the dualism could be rejected by appealing to the views on personal identity and individual rationality that Derek Parfit famously defended in his *Reasons and Persons* (1984).[23]

Among the merits of J.B. Schneewind's seminal contribution to our understanding of Sidgwick is its attention to the intellectual context in

[23] The comparison with Parfit's most recent views is pursued to some extent in Orsi's paper. Orsi critically assesses the reading of the dualism advanced by Parfit in his latest manuscript *Climbing the Mountain*.

which the *Methods* was written. By shedding light on many authors that Sidgwick discussed and referred to in his writings Schneewind greatly contributed to our understanding of the *Methods of Ethics*. The same kind of intellectual history is the focus of Massimo Reichlin and Sergio Cremaschi's papers, which examine the complex relationship of Sidgwick's thought to Kant and Whewell respectively. Massimo Reichlin draws an interesting picture of the complex web of references to Kant that can be found in Sidgwick's writings. Sidgwick had a peculiar attitude toward Kant. While explicitly mentioning him as one of his main inspiration, he never paid enough detailed attention to Kant's ethical thought. Reichlin examines some fundamental misunderstandings affecting Sidgwick's (rather scattered) references to Kant's ethical thought, and suggests that they might be due both to the influence of Mill's dismissal of Kantianism and to Sidgwick's rejection of Kant's epistemology and metaphysics.

Unlike Kant, Whewell represented a recurrent presence in Sidgwick's writings. However Sergio Cremaschi argues in his contribution that Sidgwick's treatment of Whewell is more polemical than in-depth. Sidgwick took Whewellian intuitionism to be just an abstract and generic model of conservative common sense morality. He overlooked both the specific rationalist framework developed by Whewell in his *Elements of Morality* and the detailed solutions that Whewell's texts offer to many particular moral dilemmas. Again, Sidgwick here seems to follow Mill in rejecting Whewell's ethics more on political grounds than on the basis of a careful consideration of his arguments.

Another much-debated topic in Sidgwickian scholarship is the kind of intuitionism defended in the *Methods of Ethics*. Notoriously Sidgwick grounded his justification of utilitarianism on a list of fundamental moral intuitions. Scholars however diverge about the number and the formulation of these intuitions.[24] In his piece Anthony Skelton claims that Sidgwick's utilitarianism is grounded in six fundamental intuitions and rejects rival interpretations, which generally tend to reduce the number of intuitions Sidgwick presented.[25] Skelton then goes on to show how these intuitions play a role in a complex argument for utilitarianism, which dismisses both common sense morality and dogmatic intuitionism, while presenting a

[24] See for example Rashdall 1907, pp. 90-91, 147, 184-185; McTaggart 1906; Schneewind 1977, pp. 290, 296.
[25] With the only exception of Lacey 1959.

"Millian-style" proofs of utilitarianism, where Sidgwick attempts to convince critics of utilitarianism by reliance on views that they already accept.

Today, the most pressing criticisms of utilitarianism come from virtue theorists. Sidgwick's pages anticipated also this feature of our contemporary debates. In his contribution Robert Shaver shows that, in the context of his defense of hedonism, Sidgwick's presented many different and interconnected arguments against the claim that virtue is a good (let alone the only good). This discussion appears in a chapter of the *Methods* (XIV of the book III) which Sidgwick revised many times through the various editions of his work. Shaver starts by outlining Sidgwick's main arguments and stressing the various puzzles they present. Then he argues that the best way to make sense of Sidgwick's arguments is to view them in the context of a general claim that only desirable consciousness is intrinsically good. Thus Sidgwick's argument against virtue theorists provides a way into his metaethical views of value.

Collections like the one presented in the following pages depend in fundamental ways on the generosity of their contributors. Therefore as guest editors our gratitude is mainly to them. However special thanks are also owed to Pierpaolo Marrone and the editorial board of *Etica & Politica/Ethics & Politics* for their patience and their encouragement.

Bibliography

AA. VV.
1974 *Sidgwick and Moral Philosophy*, "The Monist", vol. 58, 3, 1974.
AA. VV.
2000 *Sidgwick 2000*, "Utilitas", vol. 12, 3, 2000.
Baier, Annette C.
1985 *Postures of the Mind. Essays on Mind and Morals*, London, Methuen.
Barry, Brian
1973 *The Liberal Theory of Justice: A Critical Examination of the Principal Doctrines in a "Theory of Justice" by John Rawls*, Oxford, Oxford University Press.
Blanshard, Brand
1984 *Four Reasonable Men*, Middletown, CT, Wesleyan University Press.

Bok, Sissela
1998 *Introduction* to H. Sidgwick, *Practical Ethics. A collection of Addresses and Essays*, New York, Oxford University Press, 1998, pp. V-XIX.
Brink, David O.
1994 *Common Sense and First Principles in Sidgwick's Methods*, "Social Philosophy and Policy", 9, 1994, pp. 179-201.
Broad, Charles D.
1930 *Five Types of Ethical Theory*, London, Routledge and Kegan Paul.
Cremaschi, Sergio
2006 *Sidgwick e il progetto di un'etica scientifica*, "Etica & Politica/Ethics & Politics", 1, 2006, http://www.units.it/etica/2006_1/CREMASCHI01.htm
Crisp, Roger
2002 *Sidgwick and the Boundaries of Intuitionism*, in *Ethical Intuitionism: Re-evaluations*, Philip Stratton-Lake (ed.), Oxford, Clarendon Press, 2002, pp. 56-75.
Daurio, Janice
1997 *Sidgwick on Moral Theories and Common Sense Morality*, "History of Philosophy Quarterly", vol. 14, 4, 1997, pp. 49-67.
Gauthier, David P. (ed.)
1970 *Morality and Rational Self-Interest*, Englewood Cliffs, N.J., Prentice-Hall.
Hampshire, Stuart
1983 *Morality and Conflict*, Cambridge, Mass., Harvard University Press.
Harrison, Ross (ed.)
2001 *Henry Sidgwick*, Proceedings of the British Academy, CIX, Oxford, Oxford University Press.
Hart, Herbert L.A.
1961 *The Concept of Law*, London, Oxford University Press.
Hurka, Thomas
2003 *Moore in the Middle*, "Ethics", 113, 2003, pp. 599-628.
Lacey, Alan R.
1959 *Sidgwick's Ethical Maxims*, "Philosophy", 24, 1959, pp. 217-228.
Larmore, Charles
1987 *Patterns of Moral Complexity*, New York, Cambridge University Press.
Lecaldano, Eugenio
1996 *Henry Sidgwick. Etica teorica e razionalità pratica*, "Iride", 18, 1996, pp. 495-500.

McTaggart, John E.
1906 *The Ethics of Henry Sidgwick*, "Quarterly Review", 205, 1906, pp. 398-419.
Nussbaum, Martha C.
1986 *The Fragility of Goodness. Luck and Ethics in Greek Tragedy and Philosophy*, Cambridge, Cambridge University Press.
Pellegrino, Gianfranco
2000 *L'etica filosofica e la spiegazione del senso comune in Henry Sidgwick*, "Rivista di Filosofia", 2, pp. 309-331.
Rashdall, Hastings
1885 *Professor Sidgwick's Utilitarianism*, "Mind", o.s. 10, 1885, pp. 200-226.
1907 *The Theory of Good and Evil*, Oxford, Oxford University Press.
Rawls, John
1975 *The Independence of Moral Theory*, "Proceedings and Addresses of the American, Philosophical Association", 48, 1975, pp. 5-22.
1980 *Kantian Constructivism in Moral Theory*, "Journal of Philosophy", 77, p. 515-572, quoted from the reprint in *John Rawls: Collected Papers*, ed. Samuel Freeman, Cambridge, Mass., Harvard University Press, 1999.
1981 *Foreword to "The Methods of Ethics"*, in H. Sidgwick, *The Methods of Ethics*, Indianapolis, Ind., Hackett, pp. v-vi.
Renzo, Massimo
2008 *L'utilitarismo indiretto di Henry Sidgwick*, "Rivista di Filosofia", 3, pp. 441-466.
Schneewind, Jerome B.
1963 *First Principles and Common Sense Morality in Sidgwick's Ethics*, "Archiv für Geschichte der Philosophie", XLV, vol. 2, 1962, pp. 137-156.
1977 *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford, Clarendon Press.
Schultz, Bart
1992 (ed.) *Essays on Henry Sidgwick*, Cambridge, Cambridge University Press.
2004 *Henry Sidgwick, Eye of the Universe*, New York: Cambridge University Press.
Selby-Bigge L.A.
1890 *Some Fundamental Ethical Controversies*, "Mind", o.s. 15, 1890, pp. 93-99.

Sidgwick, Henry
1981 *Methods of Ethics*, Indianapolis, Ind., Hackett.
1996 *The Works of Henry Sidgwick*, John Slater (ed.), Thoemmes Press.
1998 *Practical Ethics. A collection of Addresses and Essays*, Sissela Bok (ed.), New York, Oxford University Press.
Singer, Marcus G.
1976 *The Methods of Justice: Reflections on Rawls*, "Journal of Value Inquiry", 10, 1976, pp. 286-316.
2000 (ed.) *Essays on Ethics and Method*, Oxford, Clarendon Press.
Singer, Peter
1974 *Sidgwick and Reflective Equilibrium*, in AA VV, 1974, pp. 490-517.
Sverdlik, Steven
1985 *Sidgwick's Methodology*, "Journal of the History of Philosophy", 58, 1985, pp. 537-553.
Taylor, Charles
1985 *Philosophy and the Human Sciences. Philosophical Papers 2*, Cambridge, Cambridge University Press.
Toulmin, Stephen
1986 *An Examination of the Place of Reason in Ethics*, 2nd ed., Chicago, Chicago University Press.
Williams, Bernard
1985 *Ethics and the Limits of Philosophy*, London, Fontana Press.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 19-41

The Dualism of the Practical Reason: Some Interpretations and Responses

Francesco Orsi
Facoltà di Filosofia
Università di Roma "La Sapienza"
francescoorsi@hotmail.com

ABSTRACT
Sidgwick's dualism of the practical reason is the idea that since egoism and utilitarianism aim both to have rational supremacy in our practical decisions, whenever they conflict there is no stronger reason to follow the dictates of either view. The dualism leaves us with a practical problem: in conflict cases, we cannot be guided by practical reason to decide what all things considered we ought to do. There is an epistemic problem as well: the conflict of egoism and utilitarianism shows that they cannot be both self-evident principles. Only the existence of a just God could, for Sidgwick, prevent the conflict and thus solve the dualism. The paper first explores in detail and rejects some reconstructions of the dualism: a purely logical account, and accounts whereby egoism and utilitarianism are principles of *pro tanto* reasons or of sufficient reasons. Then it proposes a better account, in which egoism and utilitarianism are logically compatible and yet conflicting principles of all things considered reason. The account is shown to fit with Sidgwick's view of the dualism and of its practical and epistemic pitfalls. Finally, some views are discussed as to the wider positive significance of the dualism, regarded as a challenge to the rational authority of morality, or as indicating the structural opposition of agent-relative and agent-neutral reasons, or again as the imperfect yet amendable attempt at a comprehensive pluralist theory of practical reasons.

1. *Defining the Dualism of the Practical Reason*

Henry Sidgwick famously concludes his *Methods of Ethics* (ME) with the following reasoning. The two methods, i.e., the two "rational procedure[s] by which we determine what individual human beings 'ought'…to do" (ME: 1), which have survived rational scrutiny, namely egoism and utilitarianism, can conflict in particular occasions. Mere experience shows that there is no necessary coincidence between what we ought to do on egoistic grounds and what we ought to do on utilitarian ones. The methods can conflict in this sense. Only an all powerful and just being (God), could produce a necessary coincidence, whereby, in particular, if we do what we ought on utilitarian grounds, then we do what is required by egoism: the utilitarian act will be the act that best serves our self-interest, because, being also the morally

right act, we will be rewarded for having done it by God in the afterlife (and we will be accordingly punished if we did not do it). But we cannot demonstrate, nor postulate, the existence of God, and of an afterlife. Therefore, egoism and utilitarianism can conflict, and do in fact conflict. But then there is "an ultimate and fundamental contradiction in our apparent intuitions of what is Reasonable in conduct" (ME 508). The possibility of practical conflict between egoism and utilitarianism shows that the two methods are, in some sense, contradictory. And a "contradiction" between the best methods of ethics we have looks like something we have reason to worry about.[1]

This is by and large what Sidgwick says. Admittedly, he does say something else. We have further reasons to worry about the contradiction. One is epistemological: "it would seem to follow that the apparently intuitive operation of the Practical Reason, manifested in these contradictory judgements, is after all illusory" (ibidem). This is an unclear remark. For it can be taken to mean that, when we come to contradictory judgements about what we ought to do, these very judgements grandly present themselves as the expression of Practical Reason, but, since the idea of Practical Reason expressing itself contradictorily makes no sense, we are mistaken to take either or both judgements as what Practical Reason has to say. However, this is not what Sidgwick means. Practical Reason can and does express itself contradictorily, but it should not, and that's precisely the problem. So what is illusory? If two propositions can be found to be contradictory, in themselves or in their consequences, they cannot be both intuitive, i.e. self-evident: "the propositions accepted as self-evident must be mutually consistent" (ME: 341). Hence, it is illusory to think of egoism and utilitarianism as the intuitive, self-evident expression of Practical Reason. And things are not good if Practical Reason, that "chief department of our thought", cannot issue self-evident substantive normative statements.

There is another epistemological worry. "If we gave up the hope of attaining a practical solution of this fundamental contradiction…it would [not] become reasonable for us to abandon morality altogether: but it would seem necessary to abandon the idea of rationalising it completely" (ME: 508).[2] To "rationalise morality completely" here means, more or less, to be able to find a straightforward answer to every question of what we have

[1] The expression "dualism of the practical reason" occurs in ME: xii (Preface to the Second Edition), xxi (Preface to the Sixth Edition), and 404, note 1. He regards it as "the profoundest problem of Ethics" (ME: 386, note 4).

[2] See also ME: 498.

reason to do when all things have been considered. In cases of conflict between egoism and utilitarianism, we have no straightforward answer, for — to anticipate — it will both be true and false that we ought to do a certain act. A less than complete rationalisation is, for Sidgwick, a sign of failure in our normative thought.

Finally, there is a practical worry. It is not the obvious one that in cases of conflict we just do not know what to do. Rather, it is this. Since in cases of conflict we have no more reason to do what egoism requires than to do what utilitarianism requires, whatever we do we will not have reasons and Reason on our side: "practical reason, being divided against itself, would cease to be a motive on either side; the conflict would have to be decided by the comparative preponderance of one or other of two groups of non-rational impulses" (ibidem). Practical reason ceases to be a motive in the sense that there are no further reasons to guide our decision. This does not mean that we will end up doing something for which there is no reason: self-interest or overall happiness would still provide some reason (if there were a third option which did not maximize either self-interest or general happiness, we would have no reason to choose that). But practical reason can only guide and motivate us so far. In either case we would not be able to refer to what we do as to what we ought to do period. The conflict will then *have to* be "decided" by non-rational impulses both in the sense that it would definitely be *unreasonable* to choose neither option, and that our choice of either *cannot but* represent the preponderance of one impulse over another (say, a narrow concern for our happiness, or a sense of sympathy), where there is, in the particular case, no reason for such preponderance — no matter how much we can repeat to ourselves, for instance, that acting on an utilitarian impulse is a better option because it is the morally right one.[3] Therefore this is the practical problem: Accepting the best that practical reason has to offer, i.e. egoism and utilitarianism, commits us to knowingly deliberating and acting, at least sometimes, not against practical reason, or irrationally, but *without enough* practical reason — which is puzzling, if coherent, and in practice not very comforting, especially if cases of conflict are more frequent than Sidgwick appears to believe.

Sidgwick thus is explicit — or relatively easy to interpret — o n the pitfalls of the "contradiction", but not so much on the nature of the contradiction itself. We know that it involves some contradiction between ethical judgements stemming from egoism and utilitarianism, that is in some way

[3] Strictly speaking then, it is the preponderance that is non-rational, not the impulses themselves.

generated by practical conflict, that God and only God could "solve" the conflict and prevent the contradiction. How can all of this be coherently brought together and, moreover, in such a way that we come to see Sidgwick's dualism as something to worry about, both in itself and for the reasons just discussed?

Two influential accounts of the dualism are not satisfactory, albeit for different reasons. The first is a purely theoretical account. It takes Sidgwick's talk of contradiction in its most direct sense. C. D. Broad thus expressed it:

> Sidgwick's difficulty was that *both* the principle that I ought to be *equally* concerned about equally good states of mind, no matter where they may occur, *and* the principle that I ought to be *more* concerned about a good state in my own mind than about an equally good state in any other mind, seemed to him self-evident when he inspected each separately. And yet they are plainly inconsistent with each other, so that, in one case at least an ethical principle which is in fact false must be appearing to be necessarily true (Broad 1930: 245).

This account formulates egoism and utilitarianism as mutually inconsistent theories about what we ought to be more concerned about. Utilitarianism affirms, and egoism denies, that I ought to be equally concerned about equally good states of mind no matter where they occur. It is not clear that this reading is consistent with Sidgwick's definition of methods as rational procedures for determining what we ought to do. Broad's restatement, first, does not specify a reason for being or not being equally concerned about equally good states of mind. But we would have thought that a procedure is rational insofar as it tells us what reasons determine what we ought to do. Second, egoism and utilitarianism would not be theories about what we ought to do, but about the required intensity of concern.

Moreover, if it is assumed that the object of concern are "equally good states of mind" in both cases, and one leaves "good" unspecified, then, since for Sidgwick what is good on the whole is, roughly, what anyone has reason or ought to desire, then Sidgwick's egoist, on Broad's interpretation, already ought to be concerned in some degree about others' state of mind — only, not in the same degree as hers. This may be a choice of interpretive sympathy on Broad's part: since egoism as "Pure Egoism, i.e. the doctrine that I ought not to desire to any degree as an end the occurrence of good states of

mind in anyone but myself, seems plainly false" (ibidem),[4] so the dualism would be a false problem, something we need not worry about. However, not even on Broad's construal the dualism is something we need to worry too much about: Broad's egoist already accepts that she ought to be concerned about others' happiness, and not for purely instrumental reasons. Broad's egoist, that is, is one who has embraced the "point of view of the universe". And once the egoist embraces that point of view, for Sidgwick, on the one hand, it is arbitrary to distinguish her own happiness as more important than any others' equal happiness (ME: 421); on the other hand, it is too late for her to reassert the importance of the agent's point of view, as one grounding stronger reasons for the agent. The agent's point of view — Sidgwick implies — will irreversibly cease to have its special significance. So Broad softens things up by begging the question against egoism.

But even if we amend Broad's account in these respects we will not get a fair picture of the dualism. For, by construing it as a matter of logically inconsistent principles, it makes no sense of Sidgwick's idea that, without practical conflict, and thanks to God, there would be no contradiction. Broad in fact embraces the point:

No God, however powerful and however benevolent, can alter the fact that these two principles are logically incompatible and that therefore something which seemed self-evident to Sidgwick must in fact have been false (Broad 1930: 253).

The incompatibility should have been apparent to Sidgwick from the outset. Therefore this is not a Sidgwickian interpretation of the dualism. Indeed, any purely logical account will not be Sidgwickian. One way of mending Broad's wording could be this (understanding "right" as shorthand "what there is most reason to do"):[5]

Egoism (E): There is one and only one way for an act to be right: maximizing agent-utility. Therefore acts that maximize agent-utility but not utility are right, and acts that maximize utility but not agent-utility are not right.

---

[4] If "good" here again means "good for anyone", then pure egoism would also be self-contradictory. This was G. E. Moore's (in)famous reaction to the dualism (see Moore, 1993: 150ff). I take it that this is not Broad's view (see Broad 1942), so the occurrence of "good" is an unintended slip.

[5] See Skorupski 2001: 69.

Utilitarianism (U): There is one and only one way for an act to be right: maximizing utility. Therefore acts that maximize utility but not agent-utility are right, and acts that maximize agent-utility but not utility are not right.

E and U are logically incompatible. The fact that a God could make E-right and U-right acts coincide only provides some practical reassurance. And yet for Sidgwick God could remove the contradiction itself. Thus we must make sense of a contradiction that arises not purely in virtue of the content of the principles, but also, decisively, in virtue of how the world is like: with or without God. Logical accounts of the dualism assume, uncharitably to Sidgwick, that this cannot be done.

Furthermore, it is not obvious that the logical account restores the sense of a practical problem as we construed it above.[6] Accepting the best of practical reason here would mean accepting mutually inconsistent principles: a problem for epistemic rather than practical conduct, and therefore a task for theoretical reason rather than for practical reason. Indeed, not only does the particular practical problem seen above shift off stage, but it actually disappears. Accepting E and U means accepting mutually inconsistent principles. To the extent that practical reason is constrained by theoretical reason, practical reason could hardly recommend us to act knowingly on directly logically inconsistent principles. But the practical problem stems precisely from the fact that practical reason does recommend us to act on those principles. A condition for this being the case is that accepting those principles is epistemically permissible or even feasible. Since on the logical account accepting E and U is not epistemically permissible or perhaps even feasible — they are mutually contradictory — then on the logical account practical reason does not even issue a prima facie requirement to act on either E or U. The foremost and only requirement would be to revise or reject either or both of E and U, rather than act on them. No practical question will arise before we have done our epistemic duty. require.

This is not to deny that at the heart of the dualism lies a central issue for epistemic conduct. Insofar as the two methods involve contradictory judgements, they are mutually inconsistent, and therefore the rational thing to do is revising and abandoning either or both principles with the hope of finding one (or more) that satisfy the conditions of self-evidence. But the epistemic issue should not replace the practical one: when faced with a con-

[6] On the other hand, logical accounts are well suited to explain the worries about self-evidence and complete rationalisation.

flict case, we do need to act on either egoism or utilitarianism, although there is no way of telling whether we will do the right thing. Certainly it would be wrong and unreasonable to sit back and philosophize about better principles — no matter whether this is something that we eventually shall have to do.

If the methods do not conflict purely because of their content, then we must take care to formulate such content accordingly. Here is a different, though no better, account of the dualism.[7]

E: If A maximizes agent-utility, this is a *pro tanto* reason to do A.
U: If A maximizes utility, this is a *pro tanto* reason to do A.

On the *pro tanto* account E and U are not directly incompatible. Each merely says that we have a reason *as far as* certain considerations go. In the case of an act that maximizes agent-utility, but not utility overall, it is true that we have reason to do it, and we ought to do it as far as E goes, and it is false that we have reason to do it, and we ought to do it as far as U goes. E and U pull in different directions, and we do have something worth calling a practical conflict. But so far we only have a pluralism of reasons. The dramatic, "dualist" aspect of the dualism does not appear unless we add some other proposition, such as that E and U provide incommensurable reasons.[8] If egoistic and utilitarian reasons cannot ever be weighed against each other, then, whenever they conflict, we will be unable to know what to do, and will be deceived to the extent that we think we may find a rational conclusion by weighing them. The practical problem outlined above would apply: if reason cannot postulate a minimal commensurability, we will be unassisted by reason in our final decisions.

Sidgwick may have implicitly assumed incommensurability through the metaphor of the points of view, as Derek Parfit suggests.[9] To be able to weigh egoistic and utilitarian reasons presuppose that either such reasons do not stem from different points of view, or that the two points of view are not mutually exclusive, or that there is a third all comprehensive point of view. Sidgwick does not consider either the first or the third option. Further, he denies the second one: changing point of view requires a normative *Gestalt* switch,[10] such that as egoists, we can only appreciate our and others'

[7] Skorupski 2001: 69-70.
[8] See Parfit (ms.): 113.
[9] Ibidem: 114.
[10] Skorupski 2001: 71.

egoistic reasons, and as utilitarians, we can be described as appreciating what is good for anyone, forgetting that we are not just one among others. The two points of view exclude each other, not in the sense that they directly oppose one another (this would make for logical incompatibility), but rather we are unable to inhabit both at once or somehow retain, in the switch, what we have learned to appreciate.

The *pro tanto* account, even with incommensurability, is not convincing. First, just where does the contradiction lie? If the "as far as" clause is part of the content of our judgements, these judgements are perfectly compatible all the way through. If God existed, he could prevent practical conflicts from happening, and save us from the consequences of incommensurability, but there would be no contradiction for him to remove. Hence, so this account does not explain why E and U present us with a special epistemic problem, if the source of the epistemic problem is their mutual incompatibility.

But, secondly, Sidgwick would not accept the *pro tanto* account for epistemic reasons, albeit of a different sort. As he says, any modification of apparently self-evident principles, such as to make them logically compatible, would "suggest a doubt whether the correctly qualified proposition will present itself with the same self-evidence as the simpler but inadequate one; and whether we have not mistaken for an ultimate and independent axiom one that is really derivative and subordinate" (ME: 341). This may look paradoxical: the move to *pro tanto* saves the principles from mutual inconsistence, and therefore should return them as epistemically good candidates. There is a sense, however, in which the *pro tanto* versions will not provide ultimate and independent ethical principles, but really derivative and, more importantly, *subordinate* ones. As such, they are disqualified from being self-evident, and therefore are not good enough candidates for setting up any dualism worth worrying about. (That is why we needed to add the further thought of incommensurability.) The idea is that self-evident principles must be "ultimate and independent" in their very application, i.e. they must be self-sufficient in their job of determining what is right to do, all things considered. But *pro tanto* principles are not in this sense "ultimate and independent": in order to determine all-in rightness, they depend on there not being opposing *pro tanto* principles, or reasons against doing what they favour doing. On the other hand, egoism and impartialism, taken as all things considered principles, are at least *meant* to be in this sense ultimate and independent, however they may then fail to succeed because of how the world is like. Therefore a principle lacks independence, and so self-evidence, not only if I must look "above" to see its connection to a higher principle,

but also if it implicitly directs me to look "around" in search of other possibly opposing principles.[11]

It may be that Sidgwick is misguided here, by conflating self-evidence with self-sufficiency — but we are looking for an account that makes sense of the dualism as he sees it.[12] So, guided by these last remarks, let us move on to a different one in terms of a conflict of sufficient reasons:

E: If A maximizes agent-utility, this is always a sufficient reason to do A.
U: If A maximizes utility, this is always a sufficient reason to do A.

Roughly, a sufficient normative reason to do A is a consideration which fully explains why we ought to do A all things considered. So, like all things considered reasons, and unlike *pro tanto* reasons, sufficient reasons are in the business of adjudicating each practical case. But, like *pro tanto* ones, an ultimate sufficient reason does not rule out the presence of other, possibly competing, ultimate sufficient reasons. Sufficient reasons do their explanatory job independently but not despite of each other. The move to sufficient reasons allows to make sense of the claim of egoism and utilitarianism to provide verdicts rather than just *pro tanto* reasons, without making them inconsistent with each other, i.e. without each claiming to provide the *unique* ultimate reasons. So E and U are not logically incompatible as they stand, and thus can be both self-evident. But the conflict will ensue whenever A maximizes agent-utility, but B maximizes utility, and I cannot do both. In all such cases I have sufficient reasons to do either. The problem is that weighing these reasons against each other would be worthless, however possible in principle. Since I *always* have sufficient reason to do either A or B, I am always allowed to treat either reason as the strongest one, as the one capable of deciding the case. And of course, if a further reason capable of adjudicating the case were needed, E and U would provide *in*sufficient reasons.[13] So, the only hope, again, is a powerful being that did not let conflicts arise. If God existed (with all the necessary attributes), there would necessarily be both egoistic and utilitarian sufficient reasons for the same actions. Sufficient reasons would not point in opposed directions.

---

[11] This way of reconstructing Sidgwick I partly take from Schneewind 1977: 279-80; 372-4.

[12] E.g. it doesn't seem to be a problem for David Ross's prima facie duties to be both self-evident and *pro tanto* — in this sense, not "self-sufficient".

[13] This is close to, but not exactly, Parfit's reconstruction of the dualism.

This account, however, again makes no sense of there being a real contradiction and not just an irresolvable practical conflict. Sufficient reasons determine directly what we ought to do, but each does not say: This is what and *only what* you ought to do. They do not function as excluding other possible sufficient reasons. Each claims conclusiveness for itself and not against other reasons.[14]

Moreover, if the practical conflict cannot be expressed in an inconsistent proposition, it is somehow watered down. For practical reason would seem to issue a final, *consistent*, pronouncement: Do either the egoist best act or the utilitarian best act, since there are sufficient reasons for doing either. As long as we choose either disjunct, we are doing what practical reason requires. And if we ask "Yes, but what should I do then?", it is coherent, if somewhat obnoxious, to go on answering: "Do either the egoist best act or the utilitarian best act". At this point, we will feel justified in thinking practical reason on our side all the way through, rather than only up to the point of deciding what to do. If there is sufficient reason for either disjunct, we should have no reason to worry whether we have done the right thing by choosing either. At least, this appears clear in less dramatic examples: if I have a sufficient reason for eating a chocolate icecream (the taste of chocolate) and a sufficient reason for eating a vanilla one (the taste of vanilla), and no other sufficient reasons for doing something else, and I can't eat both, then I will do the right thing whether I eat the chocolate or the vanilla icecream. I may regret having to choose (I'd rather have both) but, by definition, I do not need anything more than sufficient reasons to assure myself that I have done what is right. Sufficient reasons thus somewhat have the ability to turn the sense of conflict into a sense of comfortable choice.

2. *A Better Account*

Can a better account of the dualism be found? We have seen the conditions that need to be met: the dualism must not consist in a simple logical incompatibility, but must arise in virtue of both the content of egoism and utilitarianism, and the possibility that in a world without God the two methods conflict. Moreover, we need to state the content of the principles in such a way as to emphasize the contrast between opposing ethical perspectives.

[14] Nor does the claim that E and U *always* provide sufficient reasons make any trouble in this respect.

The first step is to formulate each principle as determining all things considered rightness or reasonableness, rather than just *pro tanto* or sufficient reasons, but without logically directly or indirectly denying each other:

(1) E: A is all things considered right if, and because, A maximizes agent-utility.
(2) U: A is all things considered right if, and because, A maximizes utility.

Plus, we need the possibility that an act may maximize, say, utility but not agent-utility and vice versa. The possibility would not arise if we could show that maximization of utility and agent-utility are necessarily inseparable, as for instance would be the case if God existed. Excluding such circumstance, Sidgwick thinks that

(3) It is possible for an act to maximize utility and not maximize agent-utility, and vice versa.

Therefore,

(4) It is possible for an act to be all things considered right and not right.

With (4) we come to see how egoism and utilitarianism lead to a genuine logical contradiction *and* how such contradiction is a consequence of facts about the principles and facts about the world. This reconstruction shows why the dualism is something to worry about both epistemically and practically. Epistemically, since E and U, plus a plausible assumption, lead to a contradiction, we should retract our judgement about the self-evidence of either or both. Indeed, leading to a contradiction is reason enough to doubt not only the self-evidence, but the very validity of either or both methods:

We cannot [...] regard as valid reasonings that lead to conflicting conclusions; and I therefore assume as a fundamental postulate of Ethics, that so far as two methods conflict, one or other of them must be modified or rejected. (ME: 6)

The worry about "complete rationalisation" is obviously explained too. If in some possible cases we are told that the same act can be right and not right all things considered, we have a contradictory answer: which is tantamount to claiming that in those cases practical reason gives us no answer. Notice the difference with the sufficient reasons account: according to that

account, in conflict cases practical reason gives us an answer, which, if unsatisfactory, is surely a meaningful and practicable one: Do either the best egoist act or the best utilitarian act. Nor is the answer unsatisfactory simply because it has a disjunctive content, but rather because we feel that this specific disjunctive content cannot always be the right answer. So the new account makes sense of the epistemic trouble that Sidgwick saw implied by the dualism.

Finally, we can give substance to the practical worry that accepting the best of practical reason leads to knowingly abandoning it (or rather to being knowingly abandoned by it) in problematic cases — when we need it most. Accepting the best of practical reason means accepting a contradiction as our guide in conflict cases. Since we cannot be knowingly guided by a contradiction, in such cases we cannot be guided by practical reason. If we choose to do either act, on the one hand we know that we are listening to *one* voice of practical reason, and we are to that extent not being wholly unreasonable; on the other hand, we know there is *another* voice of practical reason with an equal claim to be listened to, so that we cannot see ourselves as acting from such a thing as *the* verdict of practical reason.

Also, this reconstruction makes sense of the radicality Sidgwick attributes to both the egoist and the utilitarian points of view. First, it presents both principles as all things considered, i.e. having a claim to decide once and for all the normative status of every action. Second, only a further principle to the effect that, say, when the principles conflict, we should follow utilitarianism, could avoid conclusion (4). Such a principle of lexical order would imply some sort of commensurability between egoistic and utilitarian reasons. But, as we have seen, moving from the egoist to the utilitarian point of view and back again implies a normative sort of Gestalt switch, such that features like other people's well-being acquire and then lose ultimate normative relevance altogether, in a way that makes us unable to reach a stable middle ground where we can appreciate both egoist and utilitarian reasons as genuine, and therefore be in a position to compare them. That is why (4) is the conclusion of the argument. Thus we seem to have given each of Sidgwick's ingredients its due importance in our understanding of the dualism.

3. *The Responses to the Dualism*

The dualism, in the form just stated, is a philosophical embarrassment. However, the only way of getting round it that Sidgwick takes seriously is

the denial of premise (3). As Bart Schultz eloquently shows, Sidgwick's perennial interest in spiritism and telepathy reflected the need to find evidence for the possible existence of an afterlife where our self-sacrificing utilitarian efforts might be rewarded by a just and benevolent God.[15] Of course he knows well that there are other options, for instance, qualifying and therefore rejecting (1) or (2) or both as they stand. As just quoted, he assumes as a fundamental postulate of Ethics, that so far as two methods conflict, one or other of them must be modified or rejected. But he never considers any such modification, partly because the "correctly qualified proposition will [not] present itself with the same self-evidence as the simpler but inadequate one" (ME 341). Suppose both egoism and utilitarianism were qualified as *pro tanto* principles, so as to avoid the contradiction (though not necessarily the conflict). It is not obvious that they would lose anything in their apparent or real self-evidence. Nor, as we have seen above, does the suggestion that mere *pro tanto* principles would, as such, be "derivative and subordinate" (ibidem) cut any real philosophical ice, though it may be one of Sidgwick's chief reasons. More probably, the dualistic view of practical reason has here its epistemological bearing. The self-evidence of a given principle can only be appreciated by occupying the point of view relevant to the principle. Now, no weaker principles than (1) and (2) will appear self-evident when we occupy the point of view of the individual and of the universe, respectively. Since there are no other points of view to occupy insofar as practical conduct is concerned, we cannot but endorse (1) and (2) as the best that practical reason has to offer.

The last option for Sidgwick is to make sense of commensurability in order to avoid (4). But he sets things for himself in a way that precludes this. For instance, Sidgwick would have welcomed an argument showing the egoist that she is rationally required to take up the ethical "point of view of the universe". But, in Sidwgick's framework, such an argument would not work towards the commensurability of egoistic and utilitarian reasons. It is not as if we can start out as egoists and then be rationally brought to a wider perspective *while continuing to appreciate egoistical reasons as such*, so as to balance their weight against that of utilitarian reasons. When we rationally take up the point of view of the universe, egoistical considerations as such simply lose any normative weight. Any impartialist persuasion would lead us to replace our self-interested perspective with a utilitarian one, rather

[15] See Schultz 2004 on this, and in general on the development and significance of the dualism throughout Sidgwick's life.

than to expand the former into the latter. In other words, the effect of any such argument would be the complete rejection (1) in favour of (2).

What is then the proper response to Sidgwick's dualism? The answer to the question hinges on the way the dualism is understood, not only in its formal structure, as seen above, but in its philosophical significance. In this section I will not propose a response to the dualism, but rather aim at describing and evaluating some main reactions. We can fairly distinguish two major interpretive lines: (i) the dualism as presenting a general problem for normativity, and for morality in particular; (ii) the dualism as a failed attempt at constructing a comprehensive ethical view. As we will see, the two lines are not mutually exclusive.

The first title means to cover very different reactions to the dualism. What they have in common is the suggestion that Sidgwick has unveiled a deep structural or meta-ethical problem. I consider three such reactions.
David Brink argued that what is at issue is the rationality and authority of morality.[16] Recall that for Sidgwick utilitarianism is the best moral theory, in that it provides the only self-evident method for determining what is morally right and wrong. Other moral views, such as pluralist intuitionism, are shown to be defective in the self-evidence of their principles. Moreover, utilitarianism is the view that best systematizes common sense moral judgements. The dualism between utilitarianism and egoism thus is for Sidgwick coextensive with the contrast between morality itself and egoism. In conflict cases, morality and self-interest contradictorily pull in different directions.

Brink adds a further element: egoism is the best theory of rationality, just like utilitarianism is the best moral theory. If so, "the dualism of practical reason reflects the conflict between the demands of morality and those of individual rationality" (Brink 1988: 291). According to this reading, what is rationally right could be morally wrong, and what is morally right could be rationally wrong. And so we get that the same act can be all things considered right and not right. However, as Brink points out, only an externalist about morality could envisage such a dualism. Externalism is the view that "the rationality of moral considerations depends upon factors external to the concept of morality (i.e. external to the fact that the considerations in question are moral considerations). Externalism implies that it makes sense to ask whether there is reason to be moral or to do as morality requires" (ibidem: 292). On the other hand, "internalism claims that it is true in vir-

[16] See Brink 1988 and 1992.

tue of the concept of morality that moral considerations necessarily provide agents with reasons for action" (ibidem).

Now there is plenty of evidence that Sidgwick is an internalist. It is sufficient to remember that, as methods of ethics, both morality and egoism are "rational procedures": both present themselves as providing normative reasons for action. Moreover, the dualism is *of the* practical reason, that is, between two principles both belonging to practical reason, whereas on Brink's reading the dualism would be between practical reason (egoism) and something else (morality as understood by utilitarianism). So for Sidgwick moral considerations seem to be intrinsically reasonable.

While acknowledging this, Brink points out some reasons why Sidgwick might (and should) have been an externalist — and therefore why the dualism should be seen as one between morality and rationality. First, if the conflict were just one within morality, then egoism should be a plausible moral view to be set as a rival to utilitarianism. But Sidgwick hesitates to give egoism this credit, mainly because "ethical egoism seems a very implausible theory to explain and systematise our considered moral beliefs and, in particular, our beliefs about the nature and extent of our obligations to others" (ibidem: 302).

However, Brink here misses the target. An internalist reading need not conceive of the dualism as one between competing views about moral obligation.[17] Internalism takes the reasonableness of morality and egoism as given, without thereby implying that egoism is a moral view. Egoism is, rather, a view about what we ought to do from the personal point of view. So the inability of egoism to explain and systematise beliefs about moral obligations is neither here nor there. It is sufficient that egoism explains and systematises beliefs about what we ought to do from the personal or prudential point of view, i.e. when each of us considers her own existence alone, for it to count as a plausible "ethical" position to be set as a rival to utilitarianism.

Second, only an externalist reading can, for Brink, make sense of how egoism and utilitarianism conflict while being logically compatible and therefore self-evident (ibidem: 305). Egoism and utilitarianism are mutually consistent as, respectively, theories of rationality and morality. They conflict, because it is not always rational (in one's self-interest) to be moral (to act as utilitarianism requires).

Brink still assumes that on internalism egoism and utilitarianism would directly contradict each other as being both theories about morality. We

[17] Brink seems to see this but makes nothing of it (1988: 299, n.11).

know that internalists need not grant that these are conflicting theories of morality. They are better presented as conflicting theories of reasons. Brink could rejoinder that, were egoism and utilitarianism competing theories of reasons, they would be logically incompatible all the same, and therefore could not be both self-evident. In reply, however, we can point to the account offered above to show that egoism and utilitarianism can be theories of the same thing (what there is all things considered reason to do), and not be directly incompatible. And we have seen how Sidgwick's talk of a "contradiction" might be taken literally on such an account. Of course, the epistemic pitfall is that egoism and utilitarianism cannot both be self-evident. But then Brink's externalist account, if it is supposed to be preferred *because* it overcomes the self-evidence problem, begins to look like a way to solve the dualism rather than helping us understand it. The dualism is worrying, *inter alia*, precisely because it implies that egoism and utilitarianism cannot both be self-evident.

Finally, it seems that on Brink's view we lose the sense in which there is a practical conflict to be dealt with. For externalism, we may conceivably fail to have reason to do what morality requires. In this scenario, moral considerations would fail to be normative for us, just as much as the rules of etiquette might fail to be normative were there no reasons for us to follow them. But then how can moral considerations, whose normative force is contingent, conflict with an ever reason-providing egoism? It seems a platitude that only fully normative considerations can meaningfully conflict with each other. So, on the one hand, if moral considerations are not normative, there is no intelligible conflict with egoism. On the other hand, if moral considerations "acquire" normativity, then by externalism it must be in virtue of their coincidence with the results of some theory of reasons, and since egoism is the only other theory around, the dualism as Sidgwick understands it just disappears. In sum, however important Brink's problem may be, it simply is not Sidgwick's.

Parfit regards the dualism as raising a related, but different issue for morality. Here is how he formulates Sidgwick's

Dualism of Duty and Self-Interest: If duty and self-interest never conflict, we would always have most reason both to do our duty and to do what would be best for ourselves. But if we had to choose between two acts, of which one was our duty but the other would be better for ourselves, reason would give us no guidance. In such cases, we would not have stronger reasons to act in either of these ways. If we knew the relevant facts, either act would be rational. (ms.: 122)

This view is importantly different from Brink's. The intrinsic reasonableness and authority of morality is not called into question. Sidgwick is a "moral rationalist": we always have sufficient reason to do our duty or avoid acting wrongly. But, given the dualism, we cannot rule out that we might have sufficient or decisive reason to act wrongly. This would be the case every time our self-interest would be secured by a wrong action. And, to expand the thought beyond Sidgwick's views, we might have sufficient reasons to act wrongly provided by non-moral considerations of special relationships, or by what would be the impartially best outcome, in a context where this — *contra* Sidgwick — does not necessarily determine what we have strongest moral reason to do.

Sidgwick's dualism thus poses the conceptually open question: Is what we have most reason to do always morally right or permissible to do? If the answer is no, because sometimes what we have sufficient reason to do may be morally wrong, then morality is undermined in its ambition to be the supreme guide of practical reason. The point, by now familiar, is that morality, just like utilitarianism, cannot always have the last word on what we have most reason to do, because, if the dualism makes sense, at least often there is no such single last word to be had. (The term "often" is meant to reduce somewhat the extent of the dualism, as in Parfit's view of the dualism discussed below. For Sidgwick, there is never a single last word to be had in cases of conflict.)

Of course, the gravity of the problem will vary depending on what we regard as wrong. For instance, if it is held that it may on occasion be morally permissible to give priority to one's self-interest or that of one's near and dear when an impartially better outcome could be brought about, then the problem is often softened. But if it is always morally wrong to produce even an impartially slightly worse outcome by preferring a better outcome for oneself or for certain others, then it will often be the case that we have sufficient reasons to do what is wrong. Of course, since morality determines both positive and negative sufficient reasons, it will also be the case that we have sufficient reason not to do what is wrong. However, morality will only enjoy a limited authority over practical reason. We can take this to be a genuine legacy from Sidgwick's dualism.

It is worth mentioning another "structural" reading of the dualism. It is tempting to see the conflict as generated by the different kinds of reasons that become salient from the personal and universal perspectives. Personal reasons are given by facts that make reference to the agent who has them: my happiness gives me reasons to promote it, your happiness gives you rea-

sons to promote it, and so on. Impartial reasons are given by facts that make no essential reference to the agent who has them: my happiness, yours, hers… give anyone a reason to promote it as someone's happiness. The dualism would thus reflect a fundamental contrast between agent-relative and agent-neutral reasons As such, it can be expanded into a conflict between any views which countenance the same kind of thing as to be promoted (say, happiness, or perfection, or what have you) but differ as to who is to promote what, and so as to whether their reasons are agent-relative, or agent-neutral.[18] Unless we have an argument for discarding one type of reasons, there will be conflict.

This account is too thin to make sense of a deep dualism. If the turning point is the relativity or neutrality of a reason with respect to the agent, as defined above, then the dualism could be apparently solved by making all reasons agent-relative. Any agent-neutral reason-giving consideration could be stated in a way that gives every agent an agent-relative reason. For instance, one could state agent-neutral impartial reasons as self-referential altruist reasons. John's happiness, as someone's happiness, gives anyone an agent-neutral reason to promote it. But John's happiness, as *someone else's* happiness, gives anyone but John an agent-relative reason to promote it. For each agent but John, the reason-giving fact will be that "the happiness of someone else than me can be promoted". In self-referential altruism, the reason-giving fact will make ineliminable reference to the agent, albeit in a simply negative form: "x's happiness is not mine".[19] Of course, the impersonal reason each of us has to promote their own happiness merely as someone's happiness cannot be translated as a self-referential altruist reason. But naturally each agent continues to have agent-relative egoist reasons to care about his own happiness only. Now, if there are only agent-relative reasons around, it looks like the conflict will be formally solved. But such a solution would ring hollow. Having eliminated agent-neutrality and the "point of view of the universe" does not mean that we now appreciate all our reasons as stemming from our personal point of view. Surely we need to occupy the personal point of view in order to appreciate self-referential reasons, but merely occupying that point of view does not rationally commit us to appreciating all the reasons that *could* be so appreciated. The transition from egoism to pure self-referential altruism may still require a Gestalt switch even remaining within the personal point of view.

[18] See Hills 2003 for a detailed argument.

[19] See Broad 1942.

Moreover, the sense of conflict does not go. Some of the fiercest moral dilemmas arise between agent-relative reasons, such as when we would need to sacrifice our own life in order to save the life of someone to whom we are strongly attached. *A fortiori*, the sense of conflict cannot go if the sacrifice would save someone to whom we are merely related by "otherness". Confining the dualism within the personal point of view does not by any means alleviate it. Finally, these cases show that incommensurability can persist even if all reasons are agent-relative.[20] In sum, Sidgwick's dualism is best not taken to show a purely general and structural contrast between agent-relativity and agent-neutrality.

The second type of reading regards the dualism as an admirable but failed attempt at a constructive and comprehensive ethical view. According to these interpretations, Sidgwick is right insofar as he picks out two distinct and competing sources of normative reasons, but then fails to put them together in a consistent outlook, or exaggerates their incommensurability, or leaves out other sources of normativity. These theorists take their job as essentially consisting in smoothing Sidgwick over in order to come to a more reasonable and practicable view, while retaining the underlying tensions that must accompany any dualist or pluralist theory worth this name. Samuel Scheffler's "hybrid" theory (1994) makes room for agent-centred prerogatives, as grounded in the independence of the agent's perspective, to be set as limiting the moral demands of consequentialism. Likewise, Roger Crisp suggests a "dual source view" (1996) whereby *pro tanto* reasons stem both from moral requirements as given by utilitarianism and by the personal point of view. John Skorupski offers a more complex picture, whereby the dualism becomes a pluralism, as there are more ultimate sources of reasons for action than Sidgwick recognized. But among these, impartial reasons are set out as indefeasible and finally determinative of what we have overall reason to do — because they only are the expression of "pure" practical reason (2001: 78ff).[21] None of these views however really tries to deal with Sidgwick's worries.

The difficulty with Sidgwick's dualism is not only that it implies inconsistent normative statements. Taken as a normative view, it also has deeply counterintuitive consequences, as Parfit shows. In all conflict cases, we could rationally do either the best egoist act or the best utilitarian act,

[20] Parfit (ms.: 118-9) seems to believe the agent-relative/agent-neutral contrast is responsible for incommensurability or imprecise comparability.

[21] Broad's self-referential altruism (1942) can also be seen as a constructive response to Sidgwick's dualism .

whatever the strength of the relative reasons. E.g., we could rationally save ourselves from one minute of discomfort rather than saving a million people from death or agony. But "these are unacceptable conclusions. If we acted in such a way, the main reactions of others would rightly be horror and indignation. But, as well as being very wrong, our act would not be rational" (Parfit ms.: 115). This results from Sidgwick taking egoistic and utilitarian reasons to be wholly incommensurable, such that a strong impartial case in favour of an action (saving a million people from death or agony) cannot outweigh a weak egoist case in favour of a different action (saving ourselves one minute of discomfort), and vice versa. To be able to balance these reasons would mean to occupy the personal and the universal points of view at one time, and this, we know, is impossible for Sidgwick.

Parfit thinks the point of view metaphor is better discarded. As he says:

When we are trying to decide what we have most reason to do, we ought to ask this question from our actual point of view. We should not ignore some of our actual reasons merely because we would not have these reasons if we had some other, merely imagined point of view. We can also claim that, to be able to compare partial and impartial reasons, we don't need some third, neutral point of view. We can compare these two kinds of reason from our actual, personal point of view. And some reasons of either kind can be stronger than, or outweigh, some reasons of the other kind (ms.: 117).

This move also does away with the embarrassing Sidgwickian contradiction. The duality of standpoints led Sidgwick to think of each set of reasons as supreme, i.e. as determining overall rightness. But once we bring reasons together into a single point of view, each also loses such absolute aspirations, and we avoid the conclusion that the same act can be overall right and not right. At worst there will be sufficient reasons for actions that cannot be performed at the same time. But that involves no contradiction.
However, Parfit concedes to Sidgwick that all we can afford is only imprecise comparability: while different reasons are comparable, and thus each capable in principle to be stronger than another, there might be no precise truths as to their relative strength (ibidem: 113). Moreover, it may often be that the comparison, while possible, does not actually yield any unique answer as to which reason is strongest. Therefore Parfit proposes a revised version of the dualism:[22]

[22] Cp. Phillips' "indeterminacy view" of Sidgwick's dualism (1998), whereby we *never* have a determinate answer.

Wide Value-based Objective Views: When one possible act would make things go in the way that would be impartially best, but some other act would make things go best either for ourselves or for those to whom we have close ties, we *often* have sufficient reasons to act in either of these ways (ibidem: 117-8).

How often we end up with a disjunctive requirement will depend on further assumptions. As we hinted in the previous section, disjunctive requirements are not necessarily a failure of practical reason. This said, Parfit wants to leave room for situations in which the choice of either disjunct involves a deep sense of conflict, as when we have sufficient reasons to either save our own life or the life of many strangers. So Parfit's dualism is an example of what needs to be done in order to get round Sidgwick's problems while acknowledging their relative inescapability.

It is worth concluding by noting that Sidgwick himself would not have liked such a solution. Abandoning the metaphor of the opposed standpoints provides us with principles of practical reason which are both weaker than they at first sight looked, because they no longer present themselves as supreme.

Moreover, Sidgwick might doubt that impartial reasons can really be appreciated once we leave the point of view of the universe. As we have seen, some of their significance can be formally retained by viewing strangers as part of one's personal point of view, in that they are connected to oneself by some thin relation of "being other than me".

First, this particular proposal sounds paradoxical: the personal point of view, by definition, should be such that whoever and whatever is not me or mine lies beyond its normative scope. Second, even if we can make sense of others, simply as strangers or sentient beings, as lying within the personal point of view, they would be positioned at the farthest border of such a point of view. And while their relevant features, e.g. their well-being, would not for that reason count for less than those of "closer" inhabitants, it seems that, when a conflict arises between two equal distributions of well-being, the fact that in one case the benefit would be distributed among "closer" people might temptingly look like a decisive reason for us to prefer that distribution, other things being equal. In other words, if rejecting the metaphor means refusing to consider things from an imagined "the point of view of the universe", then impartial reasons risk a loss in authority which is not paralleled by a corresponding loss for personal and egoistical reasons. And

Sidgwick would have rather seen egoist reasons lose some of their authority than utilitarian ones.

Of course it might be that the actual point of view through which Parfit suggests we conduct our deliberation is not personal in any partialistic sense. But it would need to be shown why it is not so. One thought might be that, given a certain conception of personal identity, the relation one's present self has to one's future self could be as weak as, or even weaker than, the relation one's present self has to other present and future people.[23] So there would be no a priori reason to view facts about me and what is connected to me as in principle grounding stronger practical reasons than facts about other, unconnected people. The authority of impartial reasons would be no more questioned than the authority of personal and egoistical reasons. This discussion however leads us into metaphysics, and while Sidgwick would not have disliked a metaphysical solution to the dualism, it would take a different paper to explore such a possibility.

Bibliography

Brink D., 1988, "Sidgwick's Dualism of Practical Reason", *Australasian Journal of Philosophy*, 66: 291-307.
— 1992, "Sidgwick and the Rationale for Rational Egoism", in Schultz (ed.).
Broad C.D., 1930, *Five Types of Ethical Theory*, London: Routledge & Kegan Paul.
— 1942, "Certain Features in Moore's Ethical Doctrines", in P. Schilpp (ed.), *The Philosophy of G. E. Moore*, Evanston and Chicago: Northwestern University: 41-68.
Crisp R., 1990: "Sidgwick and Self-interest", *Utilitas*, 2.2: 267-80.
— 1996, "The Dualism of Practical Reason", *Proceedings of the Aristotelian Society*, 46: 53-73.
Frankena W., 1974, "Sidgwick and the Dualism of Practical Reason", *The Monist*, 58, 1974.
Harrison R. (ed.), 2001, *Henry Sidgwick*, Oxford: Proceedings of the British Academy 109.
Hills A., 2003, "The Significance of the Dualism of Practical Reason", *Utilitas*, 15.3: 315-329.

[23] The reference is obviously to Parfit's own view.

McLeod O., 2000, "What is Sidgwick's Dualism of the Practical Reason?", *Pacific Philosophical Quarterly*, 81.3: 273-90.
Moore G.E., 1993 (1903), *Principia Ethica*, Cambridge: Cambridge University Press.
Parfit D., ms., *Climbing the Mountain*.
Phillips D., 1998, "Sidgwick, Dualism and Indeterminacy", *History of Philosophy Quarterly*, 15.1: 57-78.
Scheffler S., 1994 (1982), *The Rejection of Consequentialism*, Oxford: Clarendon Press.
Schneewind J., 1977, *Sidgwick's Ethics and Victorian Moral Philosophy*, Oxford: Clarendon Press.
Schultz B. (ed.), 1992, *Essays on Henry Sidgwick*, Cambridge: Cambridge University Press.
— 2004, *Henry Sidgwick: Eye of the Universe*, Cambridge: Cambridge University Press.
Sidgwick H, 1981, *The Methods of Ethics*, 7th edition, Indianapolis: Hackett.
Skorupski J., 2001, "Three Methods and a Dualism", in Harrison (ed.): 61-81.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 42-71

Sidgwick, Origen, and the reconciliation of egoism and morality

Tim Mulgan
University of St. Andrews
tpm6@st-andrews.ac.uk

ABSTRACT
Many themes of late twentieth century ethics are prefigured in Sidgwick's *Method of Ethics*. In particular, Sidgwick's 'Dualism of Practical Reason' sets the scene for current debates over the demands of morality. Many philosophers agree that Sidgwick uncovers a deep and troubling conflict at the heart of utilitarian ethics. But Sidgwick's own response to that conflict is treated, not as a live philosophical option, but as a historical oddity. In the twenty-first century, few philosophers see the intimate connection between the dualism of practical reason and the investigation of psychic phenomena that played such a large role in Sidgwick's life. The aim of this paper is to investigate Sidgwick's own approach to the dualism of practical reason. Its general conclusion is that a non-dualistic morality demands less than a theistic religion, contrary to what Sidgwick worried - especially as concerns personal immortality and freedom.

0. *Setting the scene*

Sidgwick's *Method of Ethics* prefigures many themes of modern ethics. His 'Dualism of Practical Reason' sets the scene for current debates over the demands of morality. But Sidgwick's own solution is treated, not as a live philosophical option, but as a historical oddity. One reason for suspicion of Sidgwick's solution is its apparent affinity with traditional theism (although, as Sidgwick himself makes clear, his solution requires at most a general religious premise, and not a specifically theist one[1]). This paper resurrects Sidgwick's solution, and explores the connections and differences between the metaphysical needs of morality and those of theism. Drawing on a heretical Christian tradition going back to Origen in the third century, I argue that the metaphysical needs of theism are greater than usually supposed; while the needs of utilitarianism are much more modest.

[1] Sidgwick, *The Methods of Ethics*, p. 507, note 1. (I owe this reference to Gianfranco Pellegrino.)

1. *Sidgwick's Dilemma*

Henry Sidgwick was both the last of the great classical Utilitarians and the first modern moral philosopher. Unlike his predecessors Jeremy Bentham and J. S. Mill, Sidgwick takes moral skepticism very seriously, and asks whether morality could survive without religion. This concern is both practical (Could a secular worldview play the social role of religion?), and theoretical (Does morality even make sense in the absence of religion?) Sidgwick is less optimistic than Bentham or Mill. He believes that the decline of religion both undermines non-utilitarian moral theory, and leads to a crisis for utilitarianism.

For Sidgwick, ethics must be based on reason, not on empirical observation. Sidgwick called his masterpiece *The Methods of Ethics*. A method is a very general way of deciding what to do. Methods give rise to more specific *principles* – everyday moral rules. Sidgwick isolates three possible methods of ethics: utilitarianism, egoism, and intuitionism. For Sidgwick, the main opponents of utilitarianism are intuitionists, who believe in a "moral sense" giving us infallible knowledge of moral principles. (Sidgwick distinguishes *dogmatic* intuitionism – which he condemns – from *philosophical* intuitionism – his name for his own methodology.)

Sidgwick's first task is to demonstrate the superiority of utilitarianism to intuitionism. If I had a moral sense, I would always know what to do. As I often do not know what I ought to do, I obviously do *not* have a moral sense. Indeed, no one has a moral sense. So the intuitionist method falls apart. This leaves two competing forms of hedonism: universalistic hedonism (utilitarianism) and egoistic hedonism (egoism). These tell me to maximise the general happiness and to maximise my own happiness. Each method is an independently rational first principle. Neither takes precedence over the other. Unless the universe is specifically designed to make the two methods coincide, they will often conflict in practice. Suppose I have ten dollars. I can maximize *my own* happiness by buying a movie ticket to see *Gratuitous Violence IV*, but if I were maximizing *the general happiness* I could certainly find a better use for the money. At this point, reason offers no further guidance. Sidgwick finds an irresolvable *dualism* at the heart of human reason.

To a reader acquainted with contemporary moral philosophy, Sidgwick's dualism may seem analogous to the common objection that utilitarianism is extremely demanding.[2] However, Sidgwick himself does not explicitly

[2] For an introduction to this objection, see Mulgan, *The Demands of Consequentialism*.

worry about the *demands* of morality. Instead, he has a deeper point. His objection is not just that personal interest conflicts with the general good, or that utilitarianism is very demanding, or even that its demands are psychologically impossible. Sidgwick finds a *contradiction* in practical reason, not just a moral difficulty. Putting my own interests first is not just psychologically natural – it is also completely rational and unobjectionable. A completely selfish person commits no rational error.

For Sidgwick, the dualism of practical reason signals the failure of ethical theory. Moral philosophy must reconcile the two methods. This requirement is very strong, as contradiction is only avoided if every person's happiness always coincides exactly with the general happiness.

Sidgwick's dualism explains his enormous interest in psychic research. Individuals' interests do not coincide in the present life. Life after death is certainly not *sufficient* to solve the dualism of practical reason. The next world might be as unjust as this world. However, life after death is *necessary* for ethics. Unless there is another life where justice *might* be done, the attempt to systematise ethics is hopeless. Moral philosophers must examine the evidence that human beings can survive death. Sidgwick's paranormal activities are thus not an eccentric side-line. They are central to his philosophical concerns.

The most familiar solution combines an afterlife with God – who ensures that happiness and morality coincide. Sidgwick agrees that this solution would be satisfactory. Unfortunately, we cannot be sure that God exists. As a result, Sidgwick's own approach is more tentative. Indeed, he offers no real solution. He merely claims that any solution must involve an afterlife of some sort.

Sidgwick's own approach to his own dualism has few contemporary followers. Utilitarians ignore the possibility that we survive death, and deny that utilitarianism is incoherent if we do not survive; while religious moral philosophy is strongly anti-utilitarian. Sidgwick's problem has been much more influential in recent moral thought than his tentative solution.

2. *Why twentieth century philosophy ignored Sidgwick*

In section 3, we see how the contemporary moral philosophical landscape is moving back to Sidgwick. The present section first shows how it moved away.

The period from Moore's *Principia Ethica* in 1903 to Rawls's *A theory of justice* in 1971 was a dark age for normative ethics. The rise of philosophical

naturalism, especially in the extreme form of logical positivism, and the rejection of traditional metaphysics, undermined both Sidgwick's question and his answer.

The linguistic turn in philosophy shifts attention from normative ethics to metaethics.[3] Sidgwick's question was seldom asked. A new question became central: How do ethical facts fit into a naturalistic world view?[4] Emotivists and prescriptivists say that there are no ethical facts. Sidgwick's question thus becomes meaningless.[5] Naturalists, by contrast, identify ethical facts with natural facts. This move also undermines Sidgwick's own formulation of his dualism.

In twenty-first century philosophical vocabulary, Sidgwick is a *non-naturalist*. Ethical truth is not reducible to natural facts – not even facts about our desires. Moral philosophy seeks objective facts about what we ought to do. Such facts should be determinate. In any situation, there is only one rational thing to do. This is why the conflict between egoism and utilitarianism is so unacceptable. The gap between egoism and morality, although very troubling, is also not surprising. If ethical facts are autonomous, then there is no a priori reason to expect them to fit with our interests.

Naturalists may seem to face the same dilemma as Sidgwick. However, they need not be so troubled by it. If ethics is a matter of purely natural facts, then the failure of Sidgwick's a priori procedure is not surprising. If ethical facts are natural, then they can only be discovered a posteriori. So the naturalist can reasonably leave it to future empirical investigation to decide between egoism and utilitarianism.[6]

Even when mid-20th century moral philosophers did turn to normative ethical questions, they were often less ambitious than Sidgwick. Normative ethics offers advice, teases out the implications of alternative principles, compares theoretical approaches, and so on. The ambitious search for a single method is often replaced by a more piecemeal approach. 20th century

---

[3] Darwall et al, 'Toward *Fin de siècle* Ethics'.

[4] This question – dubbed the *location problem* by Frank Jackson – is still a central preoccupation for many moral philosophers. (Jackson, *From Metaphysics to Ethics*, chapter 5.)

[5] Although emotivists and prescriptivists reject moral facts, so do confront a conflict between prudence and morality. See, for instance, Hare, *Moral Thinking*, sections 5.5 and 6.2. (I owe this reference to Gianfranco Pellegrino.)

[6] Sidgwick himself discusses the possibility of an empirical reconciliation of prudence and morality in the concluding chapter of *The Methods of Ethics*. (I owe this reference to Gianfranco Pellegrino.)

metaethics undermined both Sidgwick's confidence in philosophical intuitionism, and his assumption that this is the only way forward for ethics.

3. *How moral philosophy is coming back to Sidgwick*

All the elements of Sidgwick's moral philosophy have made a come-back in the last few decades. The turning-point was Rawls's *A Theory of Justice* in 1971, which re-invigorated the search for ambitious, unifying theories of ethics. Non-naturalism, philosophical intuitionism, and normative ethics are firmly back on the philosophical agenda.[7] Recent analytic philosophy has also returned to the relationship between morality and religion.[8] I shall argue that the questions that have replaced Sidgwick's can benefit from answers analogous to his own.

Sidgwick sees ethics as somewhat like mathematics: a respectable autonomous realm of fact that can be explored a priori. (By contrast, logical positivists see mathematics as analytic tautology.) Many contemporary ethicists also explore connections between mathematics and ethics.[9]

The clash between egoism and utilitarianism remains a central ethical concern for contemporary utilitarian normative ethics.[10] Developments in the world beyond philosophy, such as globalisation and climate change, give Sidgwick's question a new urgency by raising new conflicts between self and others. But Sidgwick's *answer* remains ignored.

I aim to rehabilitate that answer. While Sidgwick's claims about morality and immortality are too ambitious, contemporary utilitarians can learn from them. We must first distinguish the metaphysical requirements of morality from those of theism. The metaphysical requirements of theism

---

[7] For instance, consider the theist moral realism of Robert Adams, the naturalist realism of Richard Boyd, and (especially close to Sidgwick) the non-naturalist realism of Derek Parfit. (See Adams, *Finite and Infinite Goods;* Boyd, R., 'Finite Beings, Finite Goods, Part I'; Boyd, R., 'Finite Beings, Finite Goods, Part II'; and Parfit, 'Appendix on Meta-Ethics'.)

[8] See, for instance, the recent work of Robert Adams, Linda Zagzebski, and John Bishop. (Adams, *Finite and Infinite Goods*; Zagzebski, *Divine Motivation Theory*; and Bishop, *Believing by Faith*.)

[9] See T. M. Scanlon, or Robert Adams, who harks back to Leibniz, who also regarded both mathematics and ethics as autonomous. (Scanlon, *What We Owe to Each Other*; Adams, *Finite and Infinite Goods*.)

[10] For an introduction to the current debate, see Mulgan, *The Demands of Consequentialism*, chapter one.

are seen most easily in its response to one famous objection – the argument from evil.

4. *What religion needs*

The argument from evil is central to the case against classical theism. Opponents argue that the evils of this world are inconsistent with the existence of an omnipotent, omniscient, benevolent God. In reply, the theist appeals to freedom and immortality. Evil is the price of human freedom, while an afterlife allows God to compensate the innocent victims of evil.

Theist claims about freedom and immortality can seem metaphysically extravagant. In their defence, many theists argue that morality itself makes similar claims. Theism thus involves no additional extravagance. Most famous is Kant's *moral argument*. Theoretical speculation is based on concepts designed solely for the world of experience. It cannot take us beyond that world. So it cannot tell us whether God exists, or whether we are immortal. However, morality tells me to aim for my own moral perfection and for a just world. These demands are incoherent unless their goals are possible. But they are only possible if there is an afterlife presided over by a benevolent deity. Belief in God and immortality are both practical necessities.

Sidgwick emphatically rejected Kant's argument. Given our need to systematise ethics, we have reason to *hope* that the universe is user-friendly, and a very strong motivation to seek evidence of friendliness, but this is no reason to believe that the universe actually *is* friendly. We cannot simply assume that ethics is not incoherent.

'I am so far from feeling bound to believe for purposes of practice what I see no ground for holding as a speculative truth, that I cannot even conceive the state of mind which these words seem to describe, except as a momentary, half-willful irrationality, committed in a violent access of philosophic despair.'[11]

Even if we reject the Kantian argument, a close connection between morality and religion would clearly assist theism. (Conversely, atheists may regard such a connection as an argument against morality.) I shall argue that religion and morality are not on a par. Like theism, morality does require both (a certain) freedom and (something like) immortality. But its requirements are much more modest.

---

[11] Sidgwick, *The Methods of Ethics*, Book 4, Chapter 6, p. 507.

5. *Freedom*

Theist freedom needs both a certain degree, and a certain scope. Morality requires neither that degree, nor that scope.

The free will defence presents evil as a necessary side-effect of human moral freedom. God could only avoid evil by creating automatons. Despite its evils, our world is better than any world without free agents. Contemporary philosophical debate often begins with J. L. Mackie's reply.[12] For any free agent (F) and any time (t), it is possible that F does no evil at t. It is thus *possible*, however unlikely, that F *never* does evil. The same is true of all free agents. For *any* population of free agents, there is a possible world where *those very agents* never do evil. But any perfect being will naturally choose that possible world. No perfect being will ever create a free being who ever does evil. Yet there are free beings who choose evil. Therefore, there is no God.

The now standard theist reply is due to Plantinga.[13] Plantinga does not deny that there is a possible world where free agents never do evil, nor that such a possible world is better then any where evil is done. But he denies that God could choose *that* very possible world. A free being chooses what to do without any outside determination. This is what freedom *is*. It thus makes no sense to say both that F is free, and that God chooses what F will do. Suppose the Fs are a species of genuinely free being. God can create the Fs, but God cannot choose between different possible worlds where the Fs do different things. God can only create the Fs, and then wait and see (like anyone else) what they actually do. God cannot guarantee that free agents never do evil. If free agency is sufficiently valuable, God will create free agents who might do evil. God and evil are thus not incompatible.

Plantinga requires what I call *contra divine free will* (CDF). A creature has CDF if and only if God cannot create that creature and choose its choices. Let F2 be the most valuable freedom that is not contra divine. Plantinga must claim that a world where creatures with F2 always do the right thing (w1) is worse than one where creatures with CDF sometimes do the wrong thing (w2).

---

[12] Mackie, *The Miracle of Theism*, chapter nine.

[13] Plantinga, *God and Other Minds*, chapters five and six. For a recent summary of his position, see Plantinga, *Warranted Christian Belief*, pp. 458-499.

The comparison between CDF and F2 is crucial. The creatures in w1 do not lack freedom. For all anyone knows, they may have something we would recognize as genuine freedom. In the first place, it is not obvious that every creature with libertarian freedom must also have CDF. (Given our limited understanding of the metaphysics of both libertarian freewill and divine action, we cannot be certain that God could not control the actions of a creature with libertarian freedom.) If libertarian freedom *is* logically distinct from CDF, the creatures in w1 may enjoy libertarian freedom. On the other hand, any compatibilist freedom is clearly not CDF. (If my freedom is compatible with determinism, then it is also compatible with divine control over my actions.) Therefore, if compatibilism is the correct account of human freedom, the creatures in w1 will have everything we value about our own freedom (such as moral responsibility), even without libertarian freedom.[14]

Let us concentrate on horrendous evils inflicted by one human being on another.[15] Suppose x suffers horrendous evil in w2, while no-one in w1 suffers any horrendous evil. Won't a benevolent God create w1 instead of w2, and spare x that evil?[16]

---

[14] For an introduction to the recent debate on freedom, and definitions of compatibilism and libertarianism, see Fischer et al, *Four Views on Free Will*.

[15] I borrow the term 'horrendous evil' from Marilyn Adams. (Adams, *Horrendous Evils and the Goodness of God*.)

[16] One obvious complication is Derek Parfit's *non-identity problem*. (Parfit, *Reasons and Persons*, chapter 16.) If the differences between w1 and w2 are essential to the identity of particular individuals, then w2 is not worse than w1 for anyone – as everyone in w2 would not have existed at all in w1. For ease of exposition, I put the non-identity problem to one side in the text. There are several justifications for this. First, it is obviously desirable for theism to avoid reliance on non-identity arguments, as any such defence of theism is vulnerable to attack from moral theories that can attribute moral responsibility in non-identity cases. This is especially relevant in the present case, as Parfit's original point was that utilitarian accounts cope comparatively well with non-identity situations.

Second, non-identity is very unlikely to arise for God's choices. Parfit's original argument only claims that, *as a matter of fact*, I would not have existed if things had been different. He admits that, for many of the factors that affect my identity, there is a possible world where I exist without that feature. If my parents had never met, then I *would* not exist. But there are possible worlds where I exist even though my parents never meet. (Perhaps my genetic material is brought together in a laboratory, or by magic.) We cannot bring about such worlds – but God could.

Finally, Parfit's discussion also assumes a secular account of personal identity. My identity depends (perhaps inter alia) on my genetic identity. Perhaps this account could yield a non-identity problem for God. (If my genetic makeup somehow entails that I

The freewill defence concerns the freedom to inflict horrendous evils. In w2, *this* freedom must be *contra divine*. Otherwise, God can prevent those evils. Conversely, w1 can include wide-ranging CDF – everywhere *except* when contemplating horrendous evils. Even if w1 creatures enjoy full CDF when choosing between *competing goods*, God can still ensure that w1 contains no horrendous evil.

The freedom to choose between goods is at least *as valuable* as freedom to choose between good and evil. Even if we need *some* CDF, CDF to choose evil is redundant. The additional freedom in w2 is an unnecessary – and disastrous – distraction. Indeed, choices between goods are *more* valuable. To defend this stronger claim, I present an argument that draws on the Millian utilitarian tradition, on recent work on incommensurability, on Joseph Raz's work on freedom, and my own earlier work.[17]

Our own lives include choices between competing goods, and between good and evil. We face many non-metaphysical barriers to freedom, such as sanctions, threats, or imprisonment. If these only prevent us from choosing evil over good, they do not impact on our morally valuable autonomy. Suppose I know that inflicting horrendous evil will be severely punished. This would not compromise my autonomy. Inflicting evil is not something I need to be free to do. By contrast, constraints that interfere with choices between valuable goods do reduce our well-being – sometimes quite severely. In w1, moral life centres on the choice between competing goods. w2's *only* distinctive feature is that *some* lives centre on the choice between good and evil – with some people opting for evil. The freedom enjoyed in w2 has wider scope; but this simply is not a way that w2 is superior to w1 *at all*.

This argument does not assume that autonomy has merely instrumental value. Liberal utilitarians can accord autonomy *intrinsic* value. What the argument does claim is that the intrinsic value of autonomy is found only in

---

have CDF, then I could not exist without CDF.) But alternative, non-secular, accounts make personal identity itself depend explicitly on God's will. For instance, Stephen T. Davis suggests that the fact that God wills that a certain future person is me is sufficient (in the right circumstances) to make that person my future self. (Davis, *Risen Indeed*, p. 119.) On this view, God cannot face a non-identity problem. If God says that x in possible world 1 is the same as y in possible world 2, then this makes it so. Given our uncertainty over personal identity, how could we ever know that God could not have brought it about that x, who actually has CDF, had F2 instead?

[17] Mill, *Considerations on Representative Government*; Mill, *On Liberty*; Chang, *Incommensurability, Incomparability, and Practical Reason*; Raz, *The Morality of Freedom*; Raz, "Incommensurability and Agency"; Mulgan, *Future People*.

choices between competing goods. Or, to be more precise, once we have a choice between competing goods, *then* the *addition* of a choice between good and evil does not *increase* intrinsic value. (For the purposes of the present argument, we could thus remain agnostic whether a choice between good and evil has more intrinsic value than no choice at all.) Of course, one can imagine an extreme libertarian who holds that adding the choice between good and evil does increase intrinsic value. The liberal utilitarian rejects this extreme position as intuitively implausible.

Liberal utilitarians see a shift from a focus on good and evil to a focus on competing goods as moral progress. This is not naive or optimistic. Liberal utilitarianism does not deny the role of evil in human life: it regards that role as regrettable. W1 is better for its inhabitants than W2. A benevolent God has no reason to choose W2 over w1. The horrendous evils in W2 are gratuitous.

This is an explicitly liberal utilitarian argument. It is thus not surprising that it finds support in Sidgwick's moral philosophy. Sidgwick famously defends a compatibilist account of freedom.[18] Our freedom is perfectly compatible with determinism. Sidgwick also argues that this freedom is sufficient for all moral purposes. Our lives as moral agents, our everyday decisions, and our investigations as moral philosophers require the ability to discern, weigh up, and respond to reasons. But this ability is fully compatible with our actions being ultimately determined by physical processes.

Theists typically make three claims about freedom:

1. *The Actual claim*: Human freedom is incompatibilist.
2. *The Moral claim*: Morality requires incompatibilist freedom.
3. *The Value claim*: The extra value of incompatibilist freedom outweighs the disvalue of human suffering.

Sidgwick's compatibilism rejects all three. Compatibilism itself is the denial of the actual claim. Actual evil can be justified only by our actual freedom. It is not sufficient that God might create a world containing evil. Theists must show that God might create *this world*. Sidgwick also denies the moral claim. Morality does not require incompatibilist freedom. It follows that theodicy is metaphysically more extravagant than morality. Finally, Sidgwick rejects the value claim – the heart of the free will defence. It is not sufficient that we have incompatibilist freedom, nor even that such freedom is necessary for morality. Incompatibilist freedom must also outweigh the evils of the actual world. Sidgwick is a hedonist. The only ultimate value is *"desirable consciousness"*. As a hedonist, Sidgwick places great

[18] Sidgwick, *The Methods of Ethics*, book 1, chapter 5.

value on human suffering; while, as a compatibilist, he believes that compatibilist freedom has all the value we need.

Although logically distinct, the three claims are obviously connected. Our knowledge of the value of freedom comes from introspection on our own lives and reflection on our morality. If these sources only ever deal with compatibilist freedom, then how could we know that incompatibilist freedom would be so much more valuable?

Most contemporary utilitarians follow Sidgwick's endorsement of compatibilism. They see a vast gap distance between our (morally sufficient) freedom, and what the theist needs. But the utilitarian can also convince incompatibilists, by turning to the *scope* of freedom. Morality needs freedom for three distinct purposes: to hold other people morally responsible, we must believe their actions were freely chosen; to deliberate, I must believe that my actions are under my control; and, finally, the ability to freely choose one's projects is a necessary component of a valuable human life.

Utilitarians argue that compatibilist freedom is definitely sufficient to attribute moral responsibility to others. The appropriateness of such attributions depends on the consequences of praise and blame, and involves no deeper metaphysical commitments. This is highly significant, because only the attribution of moral responsibility to others could possibly concern the freedom *to do evil*. If I am even moderately decent, then I do not seriously consider performing horrendous evils myself. So my ability to deliberate cannot depend upon my freedom to do evil. And we saw earlier that liberal utilitarians do not regard that freedom as valuable. So I have no reason to think of myself as free to do evil *at all*. Even if I must think of my freedom as CDF, I never need to ascribe evil-doing CDF *to anyone*.[19]

I conclude that morality never needs evil-doing CDF. Whatever morality does need, it needs less than theism.

6. *Immortality*

The freewill defence is typically combined with immortality. Many innocent people suffer horrendous evil without compensation – consider a young

[19] We could also note that, even if I need to think of my own freedom as *incompatibilist*, it does not necessarily follow that I must think of it as *contra divine*. For the purposes of moral deliberation, and of leading a good life, it would presumably be sufficient to believe that God could intervene in my choices, but never does.

child tortured to death. An afterlife makes compensation possible.[20] The theist then argues as follows. CDF has both benefits and costs. It makes new goods available, but it also makes horrendous evils unavoidable *by God*. The afterlife ensures that *everyone* receives the benefits, and that these benefits are sufficient to compensate for any evils suffered in this life. Ex-ante, everyone enjoys CDF plus the risk of horrendous evil. Ex post, some get CDF plus horrendous evil, while others enjoy CDF without horrendous evil. x cannot complain that *she* has suffered horrendous evil, as she benefits [perhaps post-mortem] from the features of W2 that make some evils unavoidable.

Unfortunately, an afterlife is not sufficient. Theism also needs a *prior* life. A second anti-theist argument objects, not to the *amount* of evil in the world, but to its *distribution*. Two features of that distribution are undesirable: (1) many innocent people suffer horrendous evils, while many guilty people enjoy very pleasant lives; and (2) suffering and pleasure are distributed very unequally with regard to many morally irrelevant characteristics such as gender and nationality. In short, suffering and pleasure do not track moral desert.

In a just world, suffering would not be unequally distributed in morally irrelevant ways. This does not mean there would be no suffering, but that any suffering would be distributed according to desert. Only those who deserved to suffer would do so.

If we have compatibilist freedom, or indeed any freedom other than CDF, God can ensure that no innocent person ever suffers any horrendous evil. If

---

[20] If the afterlife is infinite in duration, or contains goods of infinite value, then it may seem to completely erase horrendous evil. Suppose each finite earthly human life has a finite value. While suffering can bring this value below zero, rendering the life not worth living, it cannot create infinite disvalue. If we combine each earthly life with an afterlife of infinite positive value, then every human being enjoys an overall existence of infinite value. And, most strikingly, it seems that no amount of earthly suffering has any negative impact on that total value. By standard transfinite arithmetic, each infinite life has *the same* infinite value. Contrary to initial appearances, this world's evils do not make it worse for its inhabitants. The argument from evil collapses.
Unfortunately, this argument fails, for reasons familiar from the recent philosophical literature on infinite utility. (Vallentyne, and Kagan, 'Infinite Value and Finitely Additive Value Theory'; Mulgan, 'Transcending the Infinite Utility Debate'.) Any plausible aggregative principle for lives of infinite duration must meet the following condition: If any two lives are identical at some times, and if one is better at all times when they differ, then that life is better overall. Suppose x and y are two people who enjoy an infinitely valuable afterlife. If x's earthly life is better than y's, then x's overall existence is more valuable than y's.

we have CDF, then perhaps even God cannot prevent some innocent suffering. But God will still aim to minimise undeserved suffering. This world contains too much innocent suffering, too unequally distributed. We would not accept such unequal innocent suffering within any human society. We expect human rulers to be more impartial. We should expect no less from God. A morally perfect benevolent God would be perfectly impartial, and would not create a world where some fare so much better than others, through no merit of their own.

The best theist reply is that things are not as they seem. Imagine two otherwise identical worlds: *Rebirth* and *Single Life*. In each, many people suffer in ways that cannot be justified given their behaviour in this lifetime. The difference between the two worlds is this. In *Single Life*, each individual lives only once; while in *Rebirth*, the same individual is reborn many times, and one's fate in each life depends on one's actions in previous lives. In Rebirth, all suffering is deserved.

Rebirth is more just than Single Life. And there is no other morally significant difference – as both worlds contain the same aggregate welfare, the same average welfare, and exactly the same distribution of welfare at any one time. If desert has any value, then Rebirth is better. Any God choosing between these two worlds will prefer Rebirth.

These two possible worlds are two interpretations of our actual world. If God created the world, and if rebirth is possible, then we are living in Rebirth. There are only three possibilities: either rebirth is actual; or rebirth is logically impossible; or God does not exist. If rebirth is logically possible but not actual, then God does not exist. Theists must either defend the cycle of rebirth, or argue that it is logically impossible.

If rebirth is not possible, then God could provide a different afterlife. However, liberal utilitarians will argue instead that God would prefer not to create any human beings at all. Without rebirth, our world is simply too unjust. God would prefer creatures who never perform evil. God would create w1 instead of w2. Theism must defend the logical possibility of rebirth.

The argument that a just God would favour rebirth is not unprecedented. It can be found in all cultures where belief in rebirth is common. Nor is it unknown in the Western theist tradition – belief in reincarnation was one of the heresies attributed to Origen in the third century AD.[21] However, hav-

[21] Origen is also associated with universalism – the view that everyone (even the Devil) will eventually be saved. In fact, it seems likely that, while the accusation of universalism is just, Origen himself did not embrace reincarnation. The claim that he did is more likely to have been an attempt to discredit his views by association with aspects of con-

ing been declared a heresy, the rebirth view fell out of favour in our philosophical tradition. I argue that, in light of modern liberal utilitarian values, the time has come to reconsider that decision.[22]

7. *Is Rebirth possible?*

We begin with objections to the metaphysical coherence of rebirth (section 7), and then consider objections to a perfectly just mechanism of rebirth (section 8).

The possibility of rebirth depends on the nature of personal identity – one of the most contentious of philosophical topics. Consider two diametrically opposed positions. On a *bodily criterion*, personal identity across time requires continuity of bodily identity. It is therefore simply impossible for the same person to be reborn in different bodies. Personal survival of death requires the physical resurrection of the body – as in the traditional Christian view.[23] At the other extreme, on a *dualist* criterion, personal identity requires continuity of spiritual identity, where the soul is distinct from the body. There is then no reason why the same person cannot be reborn in different bodies. Dualism does not guarantee rebirth – or even immortality. God could simply destroy our souls at death. But dualism does mean than

---

temporary paganism considered disreputable by third-century Christians. (Edwards, *Origen Against Plato*.)

[22] Commenting on an earlier draft of this paper, Gianfranco Pellegrino raises the following problem for my argument that a cycle of rebirth could render our world just. One crucial claim in my argument is that rebirth makes it possible that seemingly undeserved suffering is actually deserved due to one's action in a previous life. Any cycle of rebirth must be either infinite or finite. Yet an infinite cycle of rebirth requires infinite past time, which is hard to reconcile with the doctrine of divine creation; while a finite cycle of rebirth implies a *first* life, where any suffering will still be undeserved. There are two main replies available to the theist. (1) If we adopt the view that God is outside time, then it may be possible for God to be the creator of a universe with an infinite past. (2) Theists could accept a first life, and argue that, as a matter of fact, there was no suffering in that life. All suffering occurred in later lives, as a result of misbehaviour in the first life. If this is a possible situation, then it must be what God has created. Nothing we observe in our lives can prove that the first life was *not* like this. (Whether they are true or not, myths of a fall from paradise are not logically incoherent.)
Finally, I would note that my dialectical purpose is to raise difficulties for theism. If the supposition that this world is just requires an infinite cycle of rebirth, and if theism is inconsistent with such a cycle, then theism is inconsistent with the supposition that this world is just.

[23] Van Inwagen, 'The Possibility of Resurrection'.

rebirth for human beings is one of God's options. Given our earlier argument, this is sufficient to establish that God would take that option.

Another currently popular view that also seems to rule out rebirth is the no-self view of Derek Parfit.[24] On this view, there is no self that continues from moment to moment. It thus seems obvious that there is no self that could survive death. We might be drawn to the no-self view by a dualist error theory. Suppose we believe that personal identity requires a soul with 'inherent existence' (in the Buddhist phrase). Finding no such soul, we conclude that there is no personal identity.

Despite appearances, Parfit's view does not automatically rule out rebirth. We must separate *eliminativism* (there are no persons) from *reductionism* (personal identity is reducible to, and no more valuable than, its constituent relations). Eliminativism rules out rebirth. But it also rejects personal identity within this life. This is very radically metaphysically revisionist. To avoid radical *moral* revisionism, eliminativists must adopt *fictionalism* about persons – for moral purposes, we talk as if there were persons, despite knowing that there are no persons. But we can then apply the same solution to rebirth. To take one striking example, even the most eliminativist Buddhist continues to speak of rebirth at the level of *conventional* truth – even while recognising the *ultimate* truth that there are no persons to be reborn.

By contrast, reductionism allows rebirth as an ultimate truth, and not merely a conventional one. Rebirth, like personal continuity within a life, can occur through memory or psychological continuity – without a separate entity that continues from one life to another. However, reductionism does create problems for our overall argument. Parfit's main point is that, because reductionism is true, personal identity is less morally significant than we are inclined to believe. If the identity of persons is nothing over-and-above certain physical or psychological relations, then it cannot be more important than those underlying relations. Reductionism leads to moral revisions, often in the direction of utilitarianism. Reductionism limits the moral significance of personal compensation and individual responsibility. It thus reduces the force of the argument from evil, and lessens the relevance of rebirth. (We return to this aspect of reductionism in the final section, where I argue that it supports our utilitarian alternative to rebirth.)

We cannot use an account of personal identity to settle the controversy over rebirth, for three reasons. The first is that personal identity is highly

[24] Parfit, *Reasons and Persons*, part three. This view is also associated with David Hume, and is found in many varieties of Buddhism.

controversial – so our account of rebirth will simply inherit that controversy. The second is that the correct account of personal identity depends upon facts about human beings. Proponents of rebirth often treat human rebirth as a datum, and thus seek an account of personal identity consistent with that 'fact'; while opponents, citing the 'datum' that humans are not reborn, may prefer a different account. Finally, our preferred account of personal identity may depend upon whether or not we believe in God. (For instance, some theists argue that the will of God can provide the mysterious 'further fact' that Parfit finds lacking in all non-reductionist accounts of personal identity.[25]) But, obviously enough, any attempt to use the resulting account of personal identity as a premise in an argument for or against the existence of God will result in circularity.

It seems that we have reached an impasse. However, we must recall our dialectical context. We are not asking whether rebirth is possible *for us*. We are asking whether there are any possible free creatures for whom rebirth is possible. If any account of personal identity consistent with rebirth is conceptually coherent, then we can *imagine* creatures for whom personal continuity is consistent with rebirth. And it seems that, whatever the truth regarding humans, dualist and reductionist accounts are coherent. Therefore, God could have created free reborn creatures. If we also believe that we are *not* such creatures, then this strengthens our objection to theism.

Consider a more modest objection to rebirth: that, whatever its conceptual coherence, rebirth is a not a plausible interpretation *of this world*. This argument appeals to the popular idea that memory is necessary for personal identity. If so, then, even if we are reborn, our rebirth typically does not preserve identity, as most people do not remember their past lives. Rebirth would then provide no personal survival beyond death. Alternatively, if we defend personal identity without memory – perhaps by appeal to an immaterial soul – we must then ask why personal identity without memory is *valuable*. [26] Can survival without memory offer compensation and punishment?

In our dialectical context, this argument against rebirth counts *against* theism. It suggests that, while logically possible, rebirth is not an epistemic

---

[25] Davis, *Risen Indeed*, p. 119.

[26] The defender of rebirth might also replace memory with *psychological* continuity – and then argue that this continuity could be *subconscious*. Perhaps my character develops through time even though I have no memory. Consider the relevance of my early years, of which I now have no memory, to my moral character. But this still leaves the evaluative questions. Is psychological continuity without memory valuable? Is it a suitable basis for desert?

possibility when applied to human beings. God could have made reborn creatures, but did not. Both theists and proponents of rebirth must reject this argument. One option is as follows. Perhaps memories of *past* lives are recovered in some *future* life. Consider the following model.[27] An individual goes through a long series of lives (L1, L2, L3, …., Ln). In the final life (Ln), all previous lives are remembered. Earlier lives are analogous to a series of dreams: each unrelated to the others, but all remembered by the single waking self. (This metaphor is especially apt within an Idealist, Buddhist, or Neoplatonic metaphysical scheme, where our final state is akin to waking from the dream of our earthly life.) The fact that some individuals do claim reliable memory of past lives is then evidence in favour of rebirth; while the fact that most people do not remember any past lives does not count against rebirth. This model seems to provide enough personal continuity to ground moral responsibility across lives. And, for all anyone knows, it is the model God has chosen.

I conclude both that rebirth is an option for a just God, and that, for all anyone knows, this is the option God has chosen. Not only might there be creatures who are reborn; but we also cannot be sure that we are not such creatures.

8. *Does rebirth guarantee justice?*

Suppose the theist concedes that rebirth is possible. They might still reject rebirth, by denying that it provides a just world. Our question was why bad things happen to good people. Rebirth offers the best reply: they do not. However, only perfectly ethicised rebirth can play this role – and this is inconsistent with CDF.

I borrow the distinction between *ethicised* and *non-ethicised* rebirth from Obeyasekere.[28] Historically, non-ethicised rebirth usually comes first. The cycle of rebirth is seem as a natural phenomenon. While it may be influenced by human action, it is not itself a moral process. In ethicised rebirth, by contrast, rebirth tracks desert. Ethicised rebirth can guarantee that people get what they deserve in the next life. Non-ethicised rebirth makes

[27] This model is drawn from McTaggart and other idealists, and is also the traditional Buddhist model of the life history of a Buddha or Arahant. (McTaggart, *Some Dogmas of Religion*; Williams, *Mahayana Buddhism*.)

[28] Obeyesekere, *Imagining Karma*.

this staggeringly unlikely. A perfectly just world requires ethicised rebirth.[29]

Suppose human beings have CDF. Suppose, also, for the sake of an argument by reductio ad absurdum, that the mechanism of rebirth is perfectly ethicised. If the rebirth mechanism is perfectly ethicised, then it must ensure that I get what I deserve in this life. My fate in this life depends, in part, on the actions of other human beings. So the rebirth mechanism must be based on perfect predictions of the actions of others. But, if such predictions are possible, then God, who is omniscient and omnipotent, could also make them. But this contradicts our assumption that humans have CDF. So the mechanism of rebirth cannot be perfectly ethicised.

Compatibilists, such as Sidgwick, will reject this argument simply by rejecting CDF. Even if we accept CDF, however, the argument still fails. CDF may rule out a perfectly ethicised system of rebirth. But *partially* ethicised rebirth mechanisms are still available. Even we, with our very limited knowledge, can make *some* predictions about an individual's fate in this world. We know, for instance, that someone born into a lower-caste family in a poor region of India has fewer life chances than someone born into affluence in the West. Presumably God can make many more predictions. The *most* just world consistent with CDF will be governed by a rebirth mechanism that is as ethicised *as possible*. Even if it is not perfectly just, this would be much more just than any world without rebirth.

Indeed, even non-ethicised rebirth might well be more just than a world where each person has only one life. If we believe in non-ethicised rebirth, then it is no longer tragic for a child to die young, as her short life is only one part of the individual's much longer journey. If every soul goes through a similar series of lives, some of them brief, then this individual's entire existence is no longer tragic *in comparison* to the total existence of others. Rebirth also allows loved ones to meet again in another life.[30] Death

[29] If the rebirth mechanism is perfectly ethicised, then we have a perfect theodicy without God. Indeed, God's only role is to act as an infallible mechanism for perfectly ethicised rebirth. If God makes *choices* independent of the individual's ethical merits, then this introduces an element of arbitrariness and unfairness.

[30] This particular role for rebirth can only be played by rebirth within the kin group, or some other system where friends in one life find each other anew in each rebirth (or at least in some future rebirth). Most systems of non-ethicised rebirth that have been adopted in human history have involved rebirth within the kin group – suggesting that, even when it is non-ethicised, one key role of belief in rebirth has always been to make the world seem more just.

thus loses much of its sting. As a result, the fact that innocent people are murdered becomes less unjust.

9. *Immortality and Morality*

We now compare the requirements of theism with those of morality. As with freedom, we distinguish both a *scope* and a *mechanism*. Theism requires a perfectly ethicised cycle of rebirth; or, if CDF makes perfection impossible, a *maximally* ethicised cycle. With regard to scope, that cycle must include all human lives – past, present, and future. A morally perfect God will create a world that not only *is* just, but *has always been* just.

It may seem obvious that morality requires much less. After all, rebirth is hardly a common view in Western culture. Many people continue to believe in morality, and to act relatively morally, without any belief in an afterlife whatsoever. The fact that belief in non-ethicised rebirth, itself insufficient for a just world, is found in many cultures reinforces the conclusion that human beings can live indefinitely within an unjust cosmos.

I agree that morality requires much less than theism in terms of both scope and mechanism. However, I shall also argue that morality does require some belief akin to immortality.

10. *Separating Morality from Theism*

I begin by dispensing with some familiar arguments that attempt to tie morality to theism. If morality requires us to believe in God, and if we cannot believe in God without an afterlife, then morality requires that afterlife. Morality might require God for three reasons. (1) If some relationship with the divine is a necessary condition for a meaningful human life, then the moral need to think of our own lives as meaningful requires belief in God. (2) Alternatively, if we can only behave morally in a world we believe to be just, and if God is necessary to guarantee justice, then we must posit God. (3) Finally, God might be necessary to ground moral truths.

All three arguments are vulnerable. Even if we agree that human lives would be *more valuable* if God existed, it does not follow that the values available in an atheist world are insufficient. Utilitarians will simply reply that the avoidance of suffering and the cultivation of the most valuable human experiences, achievements, and relationships are sufficient for a meaningful human moral life.

As we saw earlier, a perfectly just Godless world is possible, if some impersonal mechanism generates an ethicised cycle of rebirth. So justice does not require God. I would also argue that morality does not require the world to be perfectly just. We return to that question below.

Finally, God is not needed to ground moral truths. This argument for theism does have some force in relation to non-naturalists such as Sidgwick, who cannot base morality on either natural facts or human inclinations. Without God, the non-naturalist seems to leave moral facts hanging in thin air. Contemporary non-naturalists will offer two replies: one negative, the other positive. The negative reply notes that God faces the same problems as any naturalist foundation for morality – a point familiar from both Plato's Euthyphro dilemma and G. E. Moore's 'naturalistic fallacy'. Just as we can always ask, of any natural property, whether actions with that property are right; so we can ask whether God's commands are right, or the things that God loves are good. The appeal of non-naturalism rests on the implausibility of *any* foundation for moral claims, whether natural or supernatural.

The positive defence of non-naturalism would appeal to analogies with other areas of knowledge. The autonomy of different realms of discourse is a striking theme of recent philosophy. We accept knowledge of mathematics, logic, and other minds that cannot be reduced to, or derived from, knowledge of any other domain. Why not grant non-natural moral facts the same autonomy?

11. *Separating morality from justice: Scope*

Suppose we accept, as many contemporary philosophers do, that morality can survive without God. Our present question is whether it can survive without some kind of afterlife. A believer in perfectly ethicised rebirth might argue that, even if morality does not require God, it does require a perfectly just world. We saw earlier that a perfectly just world requires perfectly ethicised rebirth. So morality requires the same.

Utilitarians, like many others, will simply deny that morality presupposes a completely just world. Morality is essentially forward-looking. It relates to our impact on the world. We can affect the future, but not the past. What matters is what the future holds, not the past. So morality cannot require a belief that the world *has been* just. Indeed, utilitarians will be very suspicious of that belief. If it turns out to be false, it will have a very negative impact. (The following argument draws on a long utilitarian tradition –

especially associated with Jeremy Bentham – of rejecting conservative defences of the status quo.)

If we believe in ethicised rebirth, then we will also believe that the less fortunate deserve their misfortunes, and thus deserve no assistance. If ethicised rebirth is not true, then our false beliefs will lead us to fail to assist innocent victims of injustice. *False* belief in *ethicised* rebirth illegitimately reduces concern for the least fortunate.

If the metaphysical case for rebirth is compelling, then of course we should believe it. But the rebirth story is under-supported by evidence and argument.[31] (Even if rebirth per se is well-supported, belief in *ethicised* rebirth is certainly a leap of faith.) If we believe in rebirth, we definitely do so for moral reasons. Utilitarians will then argue that, for well-off people to believe, without sufficient evidence, that they 'deserve' their good fortune on account of virtuous past lives – while others deserve to suffer – is an extreme case of objectionable partiality.

Theism and morality have very different scopes. Theism must apply its cycle of rebirth to the past as well as the future, because a just cosmos concerns both past and future. On the other hand, for the utilitarian, morality is essentially forward-looking.

12. *Separating morality from justice: Mechanism*

Morality does not require the same scope of immortality as theism. But perhaps it requires the same mechanism, with a more limited scope. Here are three familiar moral arguments for immortality:

1. *The justice argument*. Morality tells us to play our part in making the world just. We cannot adopt a goal unless we know that goal will be achieved. Therefore, we must believe that the world will become just.
2. *The Sidgwick argument*. Morality only makes sense if there is a perfect correlation between self-interest and aggregate well-being. Such a coincidence is only possible with an afterlife. Therefore, morality requires an afterlife.
3. *The meaningfulness argument*. My life can only be meaningful if I have some chance of achieving some goal that can only be achieved if I survive death.

[31] For a philosophical critique of arguments and evidence for rebirth, see Edwards, *Reincarnation*.

All three arguments are forward-looking. But they demand different mechanisms. The justice argument is the most demanding. It requires a perfectly ethicised cycle of rebirth (or something equivalent) in the future. It is also the least persuasive argument, with two obvious weaknesses. In the first place, my goal as an individual is not a just world – something I cannot bring about – but merely to play my part in bringing about such a world. I can play that part even if I know that, because others will not play theirs, the world is unlikely to become just. Rule utilitarians have long acknowledged the distinction between an ideal code (based on an ideal world of full compliance) and moral guidance for the real world of partial compliance. The non-compliance of others is a serious moral issue, but we do not solve it by wishing it away.[32]

Furthermore, to adopt something as my goal, I clearly do not need to believe that it *will* come about. Indeed, if I already believe *that*, then it makes no sense to adopt the goal. If success is inevitable, then morality is irrelevant. The most that I must believe is that the goal is possible. If a just world is my goal, it is enough to believe that such a world is possible – however unlikely. If I have the more limited of playing my part in a just world, then I only need to believe that it is possible that my actions will make the world more just.

I conclude that the justice argument fails. We turn next to the Sidgwick argument. This requires only a correlation between self-interest and morality in the future. It does not require a just world. Indeed, the Sidgwick correlation is possible even in a world that always remains unjust. The Sidgwick correlation does not require that the good always prosper; only that some mechanism ensures that the rewards for each individual of behaving morally – whatever those rewards may be – are equal to her rewards from self-interested behaviour. It might turn out that I will suffer whatever I do, while you will prosper. What matters is that I do best by doing my duty, not that I do well.

The Sidgwick correlation clearly does not apply to this life. In this life, self-interest and morality clearly point in different directions. An afterlife can align them. So Sidgwick requires an afterlife. But he does not need an immortal afterlife, or an eternal cycle of rebirth. A perfect correlation could be achieved in a single next life – where any rewards from immorality in this life (and losses from moral behaviour) are counter-balanced.

---

[32] On contemporary rule utilitarianism, see Hooker, *Ideal Code, Real World*; Mulgan, *The Demands of Consequentialism*, chapter three; and Mulgan, *Future People*, chapter five.

So the Sidgwick correlation requires some afterlife, even if it needs much less than theism. But does morality require the Sidgwick correlation? Most contemporary utilitarians would say that it does not. Utilitarians regard the clash between self-interest and aggregate well-being as a site of real moral conflict. Our moral lives are structured by the clash between these two conflicting sources of moral demand. While it is difficult to resolve that conflict, it is not impossible. A central question for utilitarians is the extent to which morality requires me to sacrifice my own well-being for the common good. Such sacrifice is morally problematic because, in our world, it so often seems to be uncompensated. In a world with a Sidgwick correlation, while the concept of self-sacrifice may make sense, uncompensated sacrifice is ruled out.[33]

A Sidgwick correlation is not necessary for morality, and would indeed render our moral lives rather empty. Modern utilitarianism offers many more realistic ways to balance self-interest and aggregate well-being. However, although Sidgwick's correlation is unnecessary, his argument does uncover a real issue for utilitarian ethics. If the gap between self-interest and aggregate well-being grows too large, then any recognisably utilitarian moral code may become too demanding for ordinary human beings. Accordingly, utilitarians have an urgent need to seek ways to bring self-interest and aggregate well-being closer together. I shall argue that concern for future people can play this role.

We turn now to our third moral argument for immortality. The meaningfulness argument is most famously associated with Kant. Morality gives me

---

[33] This is also why utilitarians have strong reason to reject recent philosophical attempts, such as that of David Gauthier, to reduce morality to self-interested rationality. (Gauthier, *Morals by Agreement*.)

We should also note that the existence of a Sidgwick correlation would not necessarily resolve any of our *practical* difficulties. Faced with an apparent conflict between self-interest and aggregate well-being, I must decide what to do. Sidgwick tells me that the conflict is only apparent, as some unkown mechanism ensures that self-interest and aggregate well-being coincide. This does not, in itself, help me decide what to do. Should I do what self-interest seems to recommend, or what aggregate well-being seems to recommend? Commentators typically assume that the mechanism works by adjusting post-mortem individual rewards so that the action recommended by utilitarianism *in this life* also maximises self-interest. But, of course, an opposite mechanism is equally possible. Perhaps we should all pursue our own self-interest – and trust that a utilitarian afterlife will ensure that our egoism maximises aggregate well-being. (In much the same way that many of our contemporaries trust in the mechanisms of the free-market to conjure maximum aggregate well-being out of self-interest.) If our aim is to decide what to do, then positing a Sidgwick correlation does not help.

the goal of perfect virtue. As perfect virtue is only possible if I am immortal, I must adopt the postulate of immortality. As it stands, this argument is over-stated. I can surely adopt perfect virtue as the goal around which I structure my life, while still acknowledging that I cannot ever reach that goal. This move is especially congenial to utilitarians, who urge me to adopt utility maximisation as my ethical standard, without claiming that I could ever actually maximise utility.

However, like Sidgwick's dualism, Kant's argument also points to a deeper moral issue. If we are to live morally meaningful lives, then we do need something to play the role that immortality plays for Kant or Sidgwick. But a secular concern for future people is sufficient here as well.

13. *Separating future concern from self-concern*

Among both philosophers and non-philosophers, opinion divides sharply over the meaningfulness question. This division often tracks the divide between atheists and theists. Of course, atheists and theists disagree over whether they *will* survive death. But they also disagree over two evaluative questions. Would it be good to survive death? And especially: Is life meaningless or empty if we do not? Many people find it liberating to think that this life is all we have. This gives our present life a new meaning and urgency. Others find such a prospect intolerable. The former tend to be atheists; the latter theists. Of course, we could see both reactions as rationalisations. If you are convinced that this world is all there is, then you might want to look on the bright side; while someone who has devoted their life to the search for posthumous salvation will hardly cherish the prospect that this was unnecessary. But I propose to take these conflicting attitudes at face value.

My own attitude is mixed. I believe that the absence of an afterlife would not – and, indeed, *does* not – deprive life of its meaning. But, on the other hand, insofar as life is good I would like it to continue, and I certainly do not feel the force of the currently fashionable idea that an eternal life must be eventually meaningless.[34] And, most significantly for our present discussion, I believe that we must look beyond our own immediate interests and concerns – and perhaps beyond the boundaries of our own individual earthly life – to find true meaning.

[34] Williams, "The Makropolous Case".

Not everyone shares this last belief. But my present aim is to show that even those who do share it, need not posit rebirth or any other afterlife. In the meaningfulness argument, immortality plays two roles. It provides continuity of both *moral agent* and *moral object*. In Kant's original example, the two are merged, as my principal moral object *is* my own moral agency. Suppose I know that, whatever I do, I will be annihilated immediately after my next action. This fact might render my final choice meaningless, in two distinct ways. If the *objects* of my moral concern do not extend beyond my own life, then I will be indifferent to the results of my final action. Alternatively, I may feel unable to embark on any course of *action* at all – on the grounds that actions require agency extended over time, and my agency is about to end.

Drawing together our discussions of both Sidgwick and meaningfulness, I suggest that immortality can play three useful roles in moral philosophy. Immortality can provide each of the following: (1) continuity of moral agent; (2) continuity of moral object; and (3) reconciliation of self-interest and aggregate well-being.

Any form of afterlife provides for continuity of both agent and object. If I will live again, then, at any point in this life, my agency stretches into the future. Even if my moral concern is only for myself, its object is also ongoing.[35] However, no form of immortality offers a satisfactory reconciliation. Ethicised rebirth (or ethicised personal immortality) reconciles self-interest and utilitarian morality. However, because that reconciliation is perfect, it achieves too much – depriving our ethical lives of their richness and moral content.

Non-ethicised rebirth also provides continuity of both agent and object. Continuity is ensured by rebirth itself, not by its mechanism. So long as I will be reborn, both my agency and my self-concern continue. However, non-ethicised rebirth, or any other form of non-ethicised personal immortality, does nothing to reconcile self-interest and aggregate well-being. Indeed, it inhibits such reconciliation. By ensuring continuity of the *individ-*

[35] A *finite* cycle of rebirth may seem insufficient. If each life is similar, then I will face the threat of meaninglessness in my *last* life. However, if change or progress is possible from one life to another, then, while *this life* might require a next life to render it fully meaningful, we cannot assume that this requirement holds true of all future lives. Things might be different in the next life in ways that we cannot now predict, even if the next life is also finite in duration. The fact that today needs tomorrow to render today's projects meaningful, does not imply that I must live for ever if each day's projects are to make sense.

ual agent, non-ethicised rebirth allows for purely self-concerned continuity of object.

By contrast, if we reject a personal afterlife, we must also reject the possibility of continuity for the individual agent. We must then seek alternative objects of moral concern. And, as we shall now see, this very search itself inevitably leads to a partial reconciliation between self-interest and aggregate well-being.

Suppose I am convinced that I will not survive death. This threatens to make my life meaningless, especially as I approach the moment of my death. How can I ensure continuity of both moral agent and moral concern? To explore this question, let us begin with a more extreme case. Suppose I become convinced of the no-self view, advocated by Parfit and Buddhism. I see my present self, not as a continuing agent who exists through time, but as a momentarily existing atom of experience. 'I' consist only of this present choice. How can I make that choice meaningful?

If I remain self-concerned, and self-focused, my search for meaning will be fruitless. As I cease to exist the moment this choice is made, it can neither affect *me* in the future, nor form part of any meaningful ongoing pattern of action that *I* perform.

The contemporary Kantian moral philosopher Christine Korsgaard presents the need for agent continuity as a conclusive practical reason to reject the reductionist no-self account of personal identity. The fact that we can do *metaphysics* without supposing deep further facts about the identity of persons does not mean that *ethics* can afford to be equally parsimonious. To deliberate, one must see oneself as a unified conscious agent whose projects and identity endure through time.[36]

Korsgaard's Kantian argument for a continuing self is strikingly analogous to the Kantian argument for personal immortality. I shall argue that the best reductionist reply to the former provides the best utilitarian reply to the latter.

Suppose that, despite Korsgaard's argument, I remain in the metaphysical grip of the no-self view. I cannot believe in a continuing self. But I accept the need to think of my present decision as a part of some larger pattern of actions, performed by some agent larger than my (present, momentary) self. How might I proceed?

The obvious solution is to think, not in terms of individual agents, but of groups. My present self and my future selves, though not strictly one per-

[36] Korsgaard, "Personal Identity and the Unity of Agency". For related discussion, see Mulgan, "Two Parfit Puzzles" and Mulgan, *Future People*, chapter 3.

son, are still a group of agents acting in concert. Instead of focusing on *what I can do*, and then being paralysed by my own limitations, I should instead begin by asking *what we can do together*. I then choose my action, not in isolation, but because of the role it plays in some larger collective pattern of action.

The advantage of group action is that it is much more metaphysically parsimonious in this context than individual agency. I can believe that my present and future selves act as a group even if I am sceptical about the precise metaphysical status of that group. Consider the more familiar case of a group made up of different persons, such as a department or nation. I can easily believe that my department acts as a group without believing that there exists some metaphysically distinct agent that is *the department*.

This metaphysical parsimony is especially useful in the parallel case of immortality. Suppose I very strongly do not believe in personal immortality or rebirth. I simply cannot believe that I will survive death. Indeed, perhaps I cannot even entertain that belief as a 'postulate of practical reason' – whatever that means. But I am convinced that meaningfulness requires continuity of agency beyond my death. I cannot really believe that there are future selves who are continuations of my present self. But I do believe there are future people, distinct from myself, with whom I can join in group action. Instead of thinking of my present action in isolation, and despairing over the limitations imposed by my mortality, I should think of that group, and then ask what we can do together. I then play my part in our best group action.

These remarks apply to continuity of agency. This is the more difficult, and more controversial, case. Continuity of moral object is easier to achieve. Under the no-self view, as an isolated instantaneous self, I can only achieve meaning by caring about future selves who are not me. As a mortal person who rejects personal immortality, I can only achieve meaning by caring about future people who are not me. Continuity of agency obviously supports continuity of object. Once I start to evaluate my actions by considering their part in a larger group action, I am likely to begin to identify with that group, and with its other members – adopting their concerns as my own.

This brings us to a second advantage of the group action path to meaning. Unlike the solution offered by non-ethicised rebirth, it provides a partial Sidgwick correlation. To make my life meaningful, I must think more about larger wholes, and less about my individual self. This brings my self-concern closer to aggregate well-being. The reconciliation is never total. The groups in question are smaller than the whole of humanity, and my

identification with them is never absolute. The conflict between self-interest and aggregate well-being remains. But this is as it should be, if our reconciliation is not to obliterate the essence of human moral life.

Group action is hardly uncontroversial, and raises more questions than it answers.[37] But it does provide a metaphysically parsimonious alternative to both Kant's moral argument for immortality, and Sidgwick's own solution to his dualism of practical reason. It also highlights the comparative modesty of morality, as against the metaphysical extravagance of theism.

As obligations to future people become more pressing in our ethical lives, and as ethical issues become more globalised and interconnected, group actions will become ever more significant.[38] This makes our ethical lives more complicated, and can make individuals feel insignificant. I have argued that, on the contrary, group action is the key to a meaningful ethical life.

References

Adams, M., *Horrendous Evils and the Goodness of God*, Cornell University Press, 1999.

Adams, R. M., *Finite and Infinite Goods: A Framework for Ethics*, Oxford University Press, 2002.

Bishop, J., *Believing by Faith*, Oxford University Press, 2007.

Boyd, R., 'Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part I', *Philosophy and Phenomenological Research*, 66, 2003, pp. 505-553.

Boyd, R., 'Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part II', *Philosophy and Phenomenological Research*, 67, 2003, pp. 24-47.

Chang (ed.), *Incommensurability, Incomparability, and Practical Reason*, Harvard University Press, 1997.

Darwall, S., Gibbard, A., and Railton, P., 'Toward *Fin de siècle* Ethics: Some Trends', *The Philosophical Review*, 101, 1992, pp. 115-189.

Davis, S. T., *Risen Indeed: making sense of the Resurrection*, Eerdmans Publishing Company, 1993.

Edwards, M., *Origen Against Plato*, Ashgate, 2002.

[37] For an excellent recent discussion, see Woodard, *Reasons, Patterns and Cooperation*.

[38] For further discussion of these future changes in our ethical lives, see Mulgan, 'Moral Philosophy and the Future'.

Edwards, P., *Reincarnation: a critical examination*, Prometheus Books, 2002.
Fischer, J. M., Kane, R., Pereboom, D., and Vargas, M., *Four Views on Free Will*, Blackwell, 2007.
Gauthier, D., *Morals by Agreement*, Oxford University Press, 1986.
Hare, R. M., *Moral Thinking*, Oxford University Press, 1981.
Hooker, B., *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*, Oxford University Press, 2000.
Jackson, F., *From Metaphysics to Ethics*, Oxford University Press, 1999.
Korsgaard, C., 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', *Philosophy and Public Affairs*, 18, 1989, pp. 101-132.
Mackie, J. L., *The Miracle of Theism*, Oxford University Press, 1982.
McTaggart, J. McT. E, *Some Dogmas of Religion*, Edward Arnold, 1906.
Mill, J. S., *On Liberty*, first published 1859.
Mill, J. S., *Considerations on Representative Government*, first published 1861.
Mulgan, T., *The Demands of Consequentialism*, Oxford University Press, 2001.
Mulgan, T., 'Transcending the Infinite Utility Debate', *Australasian Journal of Philosophy*, 80, 2002, pp. 164-177.
Mulgan, T., 'Two Parfit Puzzles', in *The Repugnant Conclusion. Essays on Population Ethics*, J. Ryberg and R. Tannsjo (eds.), Kluwer Academic Publishers, 2004, pp. 23-45.
Mulgan, T., *Future People*, Oxford University Press, 2006.
Mulgan, T., 'Moral Philosophy and the Future', Inaugural Lecture delivered at the University of St Andrews, November, 2008.
Obeyesekere, G., *Imagining Karma: Ethical Transformation in Amerindian, Buddhist, and Greek Rebirth*, University of California Press, 2002.
Parfit, D., 'Appendix on Meta-Ethics', unpublished manuscript.
Parfit, D., *Reasons and Persons*, Oxford University Press, 1984.
Plantinga, A., *God and Other Minds*, Cornell University Press, 1967.
Plantinga, A., *Warranted Christian Belief*, Oxford University Press, 2000
Raz, 'Incommensurability and Agency', in *Engaging Reason: On the Theory of Value and Action*, Oxford University Press, 1999.
Raz, J., *The Morality of Freedom*, Oxford University Press, 1986.
Scanlon, T., *What We Owe to Each Other*, Harvard University Press, 1998.
Sidgwick, H., *The Methods of Ethics,* 7th edition, 1907.
Vallentyne, P., and Kagan, S., 'Infinite Value and Finitely Additive Value Theory', *Journal of Philosophy*, 94, 1997, pp. 5-26.

Van Inwagen, P., 'The Possibility of Resurrection' in *Immortality*, P. Edwards (ed.), Macmillan, 1992, pp. 242-246. (Reprinted from the *International Journal for Philosophy of Religion*, 9, 1978.)
Williams, B., 'The Makropolous Case: reflections on the tedium of immortality', in his *Problems of the Self*, Cambridge University Press, 1973, pp. 82-100.
Williams, P., *Mahayana Buddhism*, Routledge, 1989.
Woodard, C., *Reasons, Patterns and Cooperation*, Routledge, 2008.
Zagzebski, L., *Divine Motivation Theory*, Cambridge University Press, 2004.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 72-108

Prudence and Morality in Butler, Sidgwick, and Parfit

Alessio Vaccari
Università di Roma "La Sapienza"
alessio.vaccari@gmail.com

ABSTRACT
The debate on personal identity has profoundly modified the approach to the analysis of prudence, its structure and its links with rationality and morality. While in ethics of 18th and 19th centuries the problem of justifying prudent behaviour rationally did not exist, in contemporary ethics it seems no longer possible to justify it rationally. Particularly, from the perspective of the complex account of personal identity it seems that the only way to condemn great imprudence is from the point of view of morality. In this way we assist to a slow erosion of the clear-cut distinction between prudence and morality. The paper illustrates this change contrasting the analysis of prudence made by Joseph Butler, and then followed by his heir Henry Sidgwick, with that recently made by Derek Parfit.

1. *Introduction*

The recent debate on personal identity has profoundly modified the approach to the analysis of prudence, its structure and its links with altruism and moral theory. In contemporary thinking we witness a slow erosion of the clear-cut categorical distinction between egoism and altruism characteristic of the philosophical framework of the English 18th century[1].

In 18th century ethics, in which the problem of rationally justifying prudent behaviour did not exist, a particularly urgent need was felt to find a solution to the issue involving the possibility of accepting prudence from the specifically moral standpoint. If prudence is to be considered as a completely self-interested line of behaviour and, according to the thesis prevailing at the time, the moral point of view may be identified with an impartial and disinterested outlook, the question naturally arose of whether prudence could be reconciled, and in what way, with the need for morally virtuous behaviour.

With reference to this problem, the project of Shaftesbury and Hutcheson, in which it was claimed that a comparatively clear-cut difference existed between prudence and moral virtue, was opposed by the con-

[1] For a reconstruction of the relationship between prudence and ethics within the framework of 18th century English philosophy and the changes wrought in contemporary thinking see Eugenio Lecaldano, *L'etica e l'identità personale: tra prudenza e azione razionale*, *Archivio di filosofia*, LV n. 1-3-, pp. 231-259.

ciliatory proposal of Butler, and Smith, in which prudence was deemed to be a behaviour that did not clash with moral virtue, without however coinciding with it.

For our purpose it is important to point out that, even those who were inclined to believe a relatively strong degree of reconciliation was possible between ethical behaviour and prudence accepted the general thesis that it was possible to make a categorical distinction between these two levels of behaviour: egoistic actions motivated by self-interest ought not to be confused with altruistic or moral actions guided by benevolence.

The distinction between egoism and altruism, and consequently between prudence and ethics, albeit in a radically different philosophical context to that of utilitarianism, is not discussed in the English tradition in a book written in the late nineteenth century such as *Methods of Ethics*[2] by H. Sidgwick. And it is precisely this distinction that leads to what Sidgwick considers the most difficult problem facing ethics: the dualism of the practical reason, that is, the serious and profound contradiction between two ethical principles, that of rational egoism and that of rational benevolence, due to the simultaneous validity of two rational normative intuitions of equal weight and strength.

In contemporary philosophy we witness a shift in the axis of thinking regarding the relations between prudence and ethics versus both the 18th century framework and the theses expressed in the *Methods*. In addition to the problem of the rationality of ethics, the need is felt to raise the increasingly fundamental question of the rationality of prudence.

The turning point is represented by the huge success encountered in recent years by several ideas contained in the treatment of prudence and rational egoism given by Sidgwick in his *Methods*. In one argument, known in contemporary philosophy as the "parity argument"[3], Sidgwick describes the difficulty of considering the point of view of rational egoism fully justified and evident as compared with that of altruism. He writes:

From the point of view, indeed, of abstract philosophy, I do not see why the Egoist principle should pass unchallenged any more than the Univeralistic. I do not see why the axiom of Prudence should not be questioned, when it conflicts with present inclination, on a ground similar to that on which Egoists refuse to admit the axiom of Rational Benevolence. If the Utilitarian has to answer to the question, 'Why should I sacrifice my own happiness for

[2] Henry Sidgwick, *The Methods of Ethics* (1874, 7th ed. 1907), Indianapolis, Ind., Hackett, 1981.

[3] This term was introduced by D.O. Brink in *Sidgwick and the Rationale for Rational Egoism*, in B. Schultz (ed.), *Essays on Henry Sidgwick*, New York, Cambridge University Press, 1991, pp. 199-239.

the greater happiness of another?' it must surely be admissible to ask the Egoist, 'Why should I sacrifice a present pleasure for a greater one in the future? Why should I concern myself about my own future feelings any more than about the feelings of other persons?' It undoubtedly seems to Common Sense paradoxical to ask for a reason why one should seek one's own happiness on the whole; but I do not see how the demand can be repudiated as absurd by those who adopt the views of the estreme empirical school of psychologists, although those views are commonly supposed to have a close affinity with Egoistic Hedonism. Grant that the Ego is merely a system of coherent phenomena, that the permanent identical 'I' is not a fact but a fiction, as Hume and his followers maintain; why, then, should one part of the series of feelings into which the Ego is resolved be concerned with another part of the same series, any more than with any other series?[4]

In this well-known passage, Sidgwick emphasizes the need to provide arguments in support of prudence, and explained how the demand for these arguments derived from an analysis of the structure of prudence involving both its dimension of temporal neutrality and the implications of a complex or atomistic conception of the self. Sidgwick's approach, recently taken up again by authors like T. Nagel[5], R.M. Hare[6] and D. Parfit[7] by means of a further scrutiny of the conditions of prudent action, led to a reappraisal of the general question of the relationship between ethics and prudence.

The most significant results within this new analytical paradigm have been achieved by Derek Parfit. Following in Sidgwick's footsteps he reconstructs prudence as a theory of individual rationality, which he calls "Self-interest Theory" or S. S theory states that each agent must maximize his overall happiness, taking into consideration the probable total duration of his own life. From the content of this substantive objective it implicitly follows that each type of temporal preference must be considered irrational as the agents are asked to have an equal interest in all parts of their lives. Parfit constructs two groups of objections to the self-interest theory, reaching the conclusion that the substantive objective of this theory requires the agents to adopt an attitude vis-à-vis their own future that has no rational justification: the theory must therefore be rejected.

The outcome of these arguments thus shows that, in accordance with the classical theory of prudence, we can no longer consider imprudent actions as

---

[4] Henry Sidgwick, *The Methods of Ethics*, pp. 418-19.

[5] Thomas Nagel, *The Possibility of Altruism,* Princeton, Princeton University Press, 1970.

[6] R. M. Hare, *Moral Thinking. Its Levels, Methods and Point*, New York, Oxford University Press, 1981.

[7] Derek Parfit, *Reasons and Persons*, Oxford, Clarendon Press, 1984.

irrational since prudence does not prescribe rational actions. We therefore need a new theory that, by adopting a different criterion of rationality, will allow us to condemn imprudent actions.

Parfit believes that the only available strategy is to modify our ethical theory in such a way as to extend its application also to the class of imprudent actions which were conventionally not subject to moral evaluation.

This article thus focuses on the examination of the outcomes of the main arguments developed by Parfit to refute the "Self-interest Theory".

Furthermore, it will be attempted to demonstrate why the only valid argument among those presented in *Reasons and Persons*, is the one based on a revised conception of personal identity, thereby confirming our idea to consider Parfit's work as an essential part of the new analytical paradigm that has brought about a change in the relationship between prudence and ethics.

Before analysing the structure of the theory of self-interest and its peculiar features as a theory of rationality, I should like to examine briefly several classical treatments of prudence in the English philosophical thinking. In particular, I shall take into consideration the discussion Butler provides of "reasonable self-love" in the first few chapters of his *Fifteen Sermons Preached at Rolls Chapel* (1726)[8], as well as the more systematic treatment of "rational egoism" given by Sidgwick in Book II of his *Methods of Ethics*. In my view, these two works provide the most complete analysis of prudence in British philosophy. They show that in the past there was no problem in justifying this line of behaviour as it was deemed perfectly rational. Moreover, Butler's and Sidgwick's analysis shows that Parfit's self-interest theory has the same structural features as the classical theory of prudence and therefore to reject S amounts to rejecting the classical theory. It is of vital importance to emphasize these similarities: only in this way is it possible to claim that Parfit's arguments refute the classical theory of prudence and therefore legitimize this author's historical importance.

2. *Butler's "reasonable self-love"*

The most detailed treatment by Butler of the topic of prudence is contained in his most important work on ethics, the *Fifteen Sermons Preached at Rolls Chapel*, first published in 1726, and republished in a second edition with an important new preface in 1729.

---

[8] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel* (1726), with introduction, analyses, and notes by the Very Rev. W. R. Matthews, London, G. Bell & Dons LTD, 1967.

Butler's *Sermons* are presented as a treatment of interconnected topics and arguments mainly of ethical nature that have however often been developed unsystematically and are therefore hard to interpret.

Before making a direct examination of Butler's conception of "reasonable self-love" let us make a schematic overview of his conception of human nature, within which this doctrine is situated.

2.1. *The conception of human nature*

The importance of investigating the notion of human nature in the reconstruction of Butler's fundamental ethical theses is underlined by the author himself in the Preface to his *Sermons*:

> They were intended to explain what is meant by the nature of man, when it is said that virtue consists in following, and vice in deviating from it; and by explaining to show that the assertion is true.[9]

In the same Preface, Butler tells us that the principle according to which virtue lies in following nature is a very old one and has its origin in the ethical reflections of Stoic thinking. This principle, which the author believes still to be valid, requires a different proof from that which has been given in the past in which its psychological implications rather than its metaphysical implications are highlighted. In Sermons I, II, III and XI, Butler is engaged in the reconstruction of a conception of human nature on the basis of which virtuous behaviour is the only behaviour fully compliant with man's true constitution.

According to Butler, human nature consists of a plethora of internal principles that can easily be distinguished from each other, in spite of the philosophers' tendency to confuse them:

> Mankind has various instincts and principles of action, as brute creatures have; some leading most directly and immediately to the good of the community, and some most directly to private good.
>
> Man has several which brutes have not; particularly reflection or conscience, an approbation of some principles or actions, and disapprobation of others.[10]

[9] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Pref., p. 7.
[10] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Pref., p. 12.

The term "principle" is the most commonly noun used by Butler to refer indifferently to any internal source of human action. According to Butler's analysis, among the practical principles, it is important to make the distinction between "particular affections, appetites and passions" and "principle or general affection of self-love", and between the latter and the "natural principle" of benevolence. In addition to these elements it is possible to identify a principle of reflection which is used by mankind to express moral approval or disapproval of their actions, which Butler calls "conscience".

Self-love is identified with a general affection that urges us to act in conformity with our happiness; the general character of its object differentiates it from the other particular affections. It is also a rational principle, as it implies a capacity to distinguish between our present desires and our overall well-being. For the time being we shall not dwell on the analysis of this affection, to which the following section will be devoted.

The meaning Butler attributes to the principle of benevolence is more problematic. The status of this affection has caused a divergence among his commentators. The fundamental issue is whether this is a general principle, distinct from particular passions, or whether instead the term "benevolence" is simply a general term that refers indifferently to all the particular desires that have the welfare of others as their object[11]. It is beyond our present scope to clarify the different positions related to this problem. In agreement with T. Penelhum[12], I shall assume that the principle of benevolence must be interpreted as the "love of our neighbour". Benevolence is therefore a general desire whose object is not universal good but only the good of "that part of mankind, that part of our country, which comes under our immediate notice, acquaintance and influence, and with which we have to do".[13] Butler stresses that this general affection is also a rational principle: it actually demands the capacity to distinguish in others their short-term satisfactions from their long-term good.

The last principle identified in human nature by Butler is conscience. Unlike benevolence and self-love, conscience is not an affection. Butler presents it as "a principle of reflection in men, by which they distinguish be-

---

[11] For a detailed discussion of this problem see Terence Penelhum, *Butler*, London, Routledge & Kegan Paul, 1985. For a comparison of Butler's views on benevolence with those of Hutcheson and Hume see T. A. Roberts, *The Concept of Benevolence*, London, Macmillan, 1973; see also Amelie Rorty, *Butler on Benevolence and Conscience*, "Philosophy" 53 (1978), pp. 171-181. For a recent and extended study of Butler's ethics, see Stephen Darwall, *The British Moralists and the Internal 'Ought'*, Cambridge, Cambridge University Press, chap. 9.

[12] See Terence Penelhum, *Butler*, pp. 31-35.

[13] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. XII, 3, pp. 186-187.

tween, approve and disapprove, their own actions".[14] The language used by Butler suggests that conscience is the only faculty by means of which human beings reflect on their own nature and on the practical principles and actions that comply with it. However, if what has been stated is correct, it is obvious that also self-love and benevolence demand a certain capacity for reflection as, in order to function correctly, each requires an awareness either of one's own nature and one's own needs or of that of the others and their needs. Therefore, while it is true that conscience must have a distinctive nature, this nature must reside in that special way in which it exerts its reflective power. In Sermon I, Butler strongly emphasizes that the distinctive nature of its judgments depends on the form they take on: namely, that of approval or disapproval. Conscience is therefore that faculty by means of which we approve or disapprove of our actions and the practical principle through which we act. In this sense, conscience can bridle self-love and benevolence when they lead us to perform actions that, in spite of our intentions, are contrary to our long-term interests and those of others.

The author of the *Sermons* believes that the empirical evidence in support of the presence of these principles in human nature is sufficient *per se* to demonstrate that human beings are not only liable to neglect the interests of others, but may also neglect their own. In this sense, promoting the good of others could be in agreement with the constitution of human nature to the same extent as acting to pursue one's own personal good.

However, even if this were sufficient to demonstrate that virtuous actions are to some extent compliant with human nature, it does not amount to claiming that virtue is compliant with our nature in a way that vice is not. To support this thesis it is necessary to attribute a well-defined meaning to the notion of "following nature". Butler lists three possibilities. It may be done by acting in compliance with one of the internal principles; or else simply by following the principle that, at some particular moment in time, has greater force than the others; lastly, we may follow our nature, acting in a way that is compliant with our entire constitution. Butler believes that only in the latter meaning it is possible to defend the thesis that virtue is compliant with human nature.

What does the concept of "entire constitution" of human nature refer to?

In Butler's conception, human nature, unlike that of the other animals, is organized in such a way that its principles have a superiority over particular inclinations which differs from that deriving from the simple intensity of the motivating force: their superiority is due to their status. Acting unnaturally simply means following a particular affection that one of the superior principles under those particular circumstances would ask us not to follow.

---

[14] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. I, 8, p. 38.

Butler illustrates the point by referring to the unnaturalness of those actions we perform when we are in the sway of a violent desire, even though we know it will lead us to our doom:

Now what is it which renders such a rash action unnatural? Is it that he went against the principle of reasonable and cool self-love, considered *merely* as a part of nature? No: for if he had acted the contrary way, he would equally have gone against a principle or part of his nature, namely, passion or appetite. But to deny a present appetite, from foresight that the gratification of it would end in immediate ruin or extreme misery, is by no means an unnatural action: whereas to contradict or go against cool self-love for the sake of such gratification, is so in the instance before us. Such an action then being unnatural; and its being so not arising from a man's going against a principle or desire barely, nor in going against that principle or desire which happens for the present to be strongest … There must be some other difference or distinction to be made between these two principles, passion and cool self-love, than what I have yet taken notice of. And this difference, not being a difference in strength or degree, I call a difference in *nature* and in *kind*. And since, in the instance still before us, if passion prevails over self love, the consequent action is unnatural; but if self-love prevails over passion, the action is natural: it is manifest that self-love is in human nature a superior principle to passion. This may be contradicted without violating that nature; but the former cannot. So that, if we will act conformably to the economy of man's nature, reasonable self-love must govern. Thus … we may have a clear conception of the *superior nature* of one inward principle to another; and see that there really is this natural superiority, quite distinct from degrees of strength and prevalency.[15]

As Butler emphasizes in this passage, the question at stake in the case of the unnaturalness of serious rashness is whether to satisfy this impulse under those circumstances is contrary to the dictates of "self-love", a practical principle that, in accordance with the hierarchical structure of human nature, exerts its authority over any present inclination.

Butler claims that conscience has a supreme authority over all the other practical principles. In this sense, it may be claimed that acting against one's nature means following a lesser principle rather than the authority of conscience.

The doctrine of the natural authority of conscience is Butler's best known contribution to ethical theory, even though it is the one that caused his interpreters the greatest difficulty. In particular, it is not clear to what extent

[15] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. II, 11, pp. 55-56.

the superiority of the conscience demands behaviour that may clash with the prescriptions of self love. In several passages, indeed, Butler seems to accept the thesis that moral virtue coincides with our own interests, if not actually to defend the much stronger thesis that to follow the dictates of conscience is justified solely by the fact that self love prescribes the same course of action.

2.2. *"Reasonable self love" and particular affections*

Before going on to analyse the specific nature of the self-love principle and the differences between it and particular affections, it is necessary to make a few further considerations regarding Butler's moral psychology.

Butler claims that to have an affection means having a particular goal, and that such a goal must be considered the object of that affection:

> Now, as reason tends to and rests in the discernment of truth, the object of it; so the very nature of affection consists in tending towards, and resting in, its objects as an end. … If we have no affections which rest in what are called their objects, then what is called affection, love, desire, hope, in human nature, is only an uneasiness in being at rest; an unquiet disposition to action, progress, pursuit, without end or meaning.[16]

The notion of the object of an affection thus finds its primary use in those cases in which the affection concerned has a particular aim or objective. The paradigmatic cases are presumably those in which the latter is a desire, of which appetites are special cases[17]. Other possible examples could be those of passions that are not desires, such as anger, resentment or compassion, in which however the psychological state under scrutiny is itself logically linked to the desire to do good or evil to someone.

The notion of the object of an affection allows us to appreciate the difference between particular impulses and self love. In his Preface to *Sermons*, Butler introduces the self-love principle, contrasting it with affections. The true contrast, however, is not with affections as such but with the particular nature of their objects. Self love is a general principle, not a special one; its general nature derives from its specific object, namely, happiness, or the long-term good of the subject. Butler writes:

---

[16] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. XIII, 5, pp. 206-7.

[17] According to Butler's terminology, appetites are desires related to physical survival or well-being: such as hunger, thirst or sexual desire.

[…] private happiness or good is all which self-love can make us desire, or be concerned about: in having this consists its gratification; it is an affection to ourselves; a regard to our own interest, happiness and private good: and in the proportion a man hath this, he is interested, or a lover of himself […].[18]

Butler does not dwell at length on discussing what happiness represents for human beings; this is probably due to the general nature of self love, which excludes too precise an understanding of its object.

Happiness or satisfaction consists only in the enjoyment of those objects, which are by nature suited to our several particular appetites, passions and affections.[19]

Whatever the particular choices of each person that lead to happiness, a happy life will be one in which the majority of particular appetites and passions are satisfied in the long term.

The partial characterization of happiness provided by Butler helps clarify the status of self love. As several scholars have emphasized, the latter is a second order practical principle: that is, it is a desire to satisfy several desires and for certain objects of aversion to be removed in the long term.[20]

According to Butler, from this characterization it certainly emerges that self love is an affection; indeed it is a desire, although it may also be inferred that it is a rational principle. There are two reasons for this. First, self love demands the capacity to distinguish between a second order general object, such as the overall happiness of the individual, and the particular objects of passions. Second, since it may be exercised only through the judgment that particular objects will contribute to a certain extent to the general object of self love, it implies both the capacity to predict the effects of satisfying the various desires and the ability to calculate and compare the hedonic intensity associated with this satisfaction. Butler stresses the rational dimension of this principle by calling it "calm self love" and "reasonable self love". The rationality of prudent behaviour is highlighted also by the adjectives Butler uses to criticize those who are unable to achieve it: madness is often attributed precisely to those persons who are unable to master the complexity of their temporal existence, that is, who fail to compare several different satisfactions in separate times among themselves.

Butler also claims that it is precisely from the difficulty involved in achieving this line of behaviour that it follows that prudence is a virtue. In-

[18] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. XI, 8, p. 169.
[19] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. XI, 9, p. 170.
[20] T. Penelhum, *Butler*, chap. 1.

deed, insofar as it represents that reasonable self love, whose goal is our interest, prudence certainly does not coincide with immediate and particular passions. It requires the capacity to achieve proper behaviour in temporal matters: that is, to assess the consequences of one's actions beyond their more immediate effects, avoiding all actions based solely on the attainment of momentary satisfaction and, in Butler's words, not to fall into the error of forgoing a greater temporal good for a lesser one, that is, forgoing what is in our overall interest for the sake of momentary satisfaction.

2.3. *"Reasonable self love" and psychological egoism*

I want now to subject to further analysis the peculiar characteristics of the conception of prudence or the "reasonable self love" described by Butler, highlighting the differences between it and the doctrine of psychological egoism.

Psychological egoism is a thesis asserting that all actions are motivated by personal interest[21]. Two characteristics emerge from this highly schematic and general presentation. Psychological egoism is a thesis describing human nature which sets out to be empirically informative. Furthermore, this thesis is universal since it is aimed at characterizing the motives of all agents in all circumstances of action. It follows from the conjunction of these two characteristics that the conception in question might be refuted empirically.

Butler interprets the conception of psychological egoism using the language of his moral psychology: on the basis of his description, this conception asserts that all actions are performed under the influence of the affection of self love, that is, all motives can be reduced to the pursuit of one's own happiness. Such a theory has dangerous moral implications as it denies the existence of genuinely altruistic motivations. In his Sermon XI Butler will demonstrate how this doctrine must be rejected; it actually fails to stand up to the test of experience.

The first objection to psychological egoism asserts that this conception may be defended only on the basis of an erroneous interpretation of the genuine doctrine of self love. Acting in accordance with "reasonable self

[21] Psychological hedonism may be considered a special case of this thesis; it reduces all our motives to a desire for pleasure. In Sermon XI Butler devotes ample space to refuting this conception. It is due to two confused inferences: (1) the first infers from the fact that all my motives are *mine* the conclusion that the object of my desires must always be an internal state of mine, (ii) the second infers from the fact that I usually obtain pleasure from the satisfaction of my desires that my desires are always directed towards my own pleasure.

love", as already emphasized previously, means pursuing those lines of action the effects of which will presumably contribute to the overall happiness of the person. However, this in no way implies that the actions are due directly to the desire for one's own happiness. As mentioned earlier, the desire for happiness or personal well-being is a regulatory or second-order desire, the task of which is to ensure the best possible mix of satisfaction of one's particular passions. The normal function of this desire thus seems to be to approve or disapprove of particular desires in proportion to their capacity to contribute to the agent's overall happiness. It follows that this general desire cannot be a substitute for the function performed by the specific desires that are the direct cause of the actions. It seems quite apparent that, on the basis of this explanation of how self love can and should function, it is possible to come up with an immediate refutation of the doctrine of psychological egoism. Butler infers two different critiques from this.

First, it is possible to reject the thesis that our sole motivation is always the pursuit of happiness as a whole. Indeed on the basis of the explanation given above it follows that, in those circumstances in which we act in a fashion compliant with self love, there is an additional motive that the pursuit of one's own happiness encourages or at least allows to be satisfied. Butler points out how, in certain particular circumstances, it is actually possible that the pursuit of our own good is the only motive. On those occasions on which we refrain from doing something we actually desire because, for example, we deem it to be contrary to our interests, Butler asserts that self love is probably the only principle acting. However, these circumstances are rare and cannot be used as a suitable foundation for a universal theory of human motivation. Second, it is possible to reject the assumption, shared by many forms of psychological egoism, according to which the self love doctrine is incompatible with the existence of altruistic actions motivated by desires for the good of others. The essence of the discussion again hinges on the possibility of distinguishing self love and particular passions on the basis of their objects. As we saw, Butler claims that the objects specific to particular passions are particular desired objects or events, such as for example my lunch, a home, a holiday, etc. According to this distinction, he claims that although we might not be able to pursue happiness without first deciding whether our particular desires contribute to its attainment, we could act under the impulse of a particular desire without considering whether it contributes to our happiness. Once this corollary has been accepted there is no contradiction in imagining that the particular desires may include desires that contribute indirectly to the well-being of others, or actually have the good of others as their specific object. Whether these desires contribute to our happiness is a question that may be answered only through the very exercise of self love and to which no a priori answer may be given. But we can

thus conclude that desires for the good of others are not in principle more incompatible with self love than any other particular desire.

Butler's second argument is based on the observation that the substantive objective of self love, namely happiness, is attained more easily if we do not act under its impulse. He points out that in many circumstances we are better able to satisfy our happiness by not considering whether and to what extent what we wish to do contributes to it since the calculation of one's advantages can be an impediment precisely to those activities demanded by happiness:

Disengagement is absolutely necessary to enjoyment: and a person may have so steady and fixed an eye upon his own interest , whatever he places it in, as may hinder him from *attending* to many gratifications within his reach, which others have their minds *free* and *open* to.[22]

In this passage Butler points out that enjoyment is a form of attention and reflecting on self love may be a distraction. This is how he explains the fact that the persons who are always busy calculating their happiness are often unhappier than the others.

This refutation of psychological egoism has the merit of highlighting several important aspects of the theory of "reasonable self love". It actually shows not only that altruistic desires are not incompatible with the pursuit of one's own happiness but, more in general, sheds light on the risks linked to the adoption of a constant inclination to calculate the different satisfactions involved, which could lead to less satisfactory outcomes from the point of view of overall personal happiness itself.

3. *"Egoistic hedonism" in Henry Sidgwick's* Methods of Ethics

The idea that "calm self love" must be considered a normative principle, as it prescribes that individuals should pursue their overall happiness by prohibiting the satisfaction of those present inclinations which may prove detrimental in the long term, is picked up again by Henry Sidgwick in his *Methods of Ethics*. He includes "rational egoism"[23] among the methods of ethics, that is, among the rational procedures used by human beings to govern their behaviour whenever they seek to work out a complete synthesis of practical maxims.

---

[22] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, Serm. XI, 9, p. 171.

[23] Following Sidgwick's use, the terms "egoistic hedonism" and "rational egoism" will be considered synonymous.

The inclusion of "rational egoism" among the methods of ethics will lead to the failure of the foundationalist project of ethics which represented one of the principal objectives of the *Methods*. In the well-known chapter devoted to philosophical intuitionism, in which the cognitive intuitionist epistemological framework forming the background to his treatment is outlined, Sidgwick shows that two mutually incompatible principles underpin prudence and benevolence, both of which are however self-evident: consequently the egoist could coherently maintain his own position without it being possible to refute it rationally.

Before presenting several of the characteristics of rational egoism, a few general considerations will be made concerning the philosophical project of the *Methods*.

3.1. *The objectives of the* Methods of Ethics

By the expression 'method of ethics' Sidgwick means any rational procedure by means of which it is possible to determine what human beings as single individuals must do. One of his strongest convictions is that common sense morality embodies different methods:

> Still I think that when a man seriously asks 'why he should do' anything, he commonly assumes in himself a determination to pursue whatever conduct may be shown by argument to be reasonable [...] And we are generally agreed that reasonable conduct in any case has to be determined on principles [...] But when we ask what these principles are, the diversity of answers which we find manifestly declared in the systems and fundamental formulae of professed moralists seems to be really present in the common practical reasoning of men generally [...].[24]

In Chapter 1 of Book 1, Sidgwick briefly discusses a variety of methods and principles which are linked in different ways and through different factual assumptions. More precisely, as J.B. Schneewind[25] emphasizes, Sidgwick, by analysing human moral reasoning, had identified two types of methods of ethics: 1) methods logically linked to the ultimate principles, and 2) methods indirectly linked to the ultimate principles. A method logically linked to an ultimate principle requires the moral agent to identify the

---

[24] Henry Sidgwick, *The Methods of Ethics*, p. 6.

[25] Jerome B. Schneewind, *Sidgwick Ethics and Victorian Moral Philosophy*, Oxford, Clarendon Press, 1977, cap. 6, pp. 194-98. See also, J. Schneewind, *Sidgwick and the Cambridge Moralists*, in Bart Schultz (ed.), *Essay on Henry Sidgwick*, pp. 93-121.

action to be performed exclusively through the only property that renders the actions right (right-making property). On the other hand, in a method indirectly linked to the ultimate principle, the moral agent identifies the actions to be performed not through the sole right-making property but by means of a characteristic linked to the latter through a contingent link (a criterial property). As Sidgwick himself asserts, his treatment was concerned solely with the "critical exposition of the different 'methods' … which are logically connected with the different ultimate reasons widely accepted".[26] The reason for this restriction is probably to be sought in the fact that Sidgwick was aware that one of the main causes of disagreement among human beings concerning their specific moral judgments consists of the differences related to their psychological, religious or metaphysical beliefs. By insisting on this restriction the moral philosopher was able to eliminate all the difficulties pertaining to realms of thinking that lay beyond the scope of ethics to investigate.

Among the many methods that are cloaked in varying degrees in the ambiguity of our moral language, Sidgwick claims that the following three methods can be distinguished: "egoistic hedonism", universalistic hedonism or utilitarianism, and intuitionism. He asserts the widely accepted commonsense view that it is rational to act both for one's private happiness as a whole and for the general happiness of all individuals. In this way it is easy to generate both the method of egoism and that of utilitarianism.

The intuitionist method, unlike the other two, is not linked directly to an ultimate principle. For the sake of simplicity intuitionism could be defined as the theory of ethics which considers as the ultimate aim of moral actions their compliance with certain unconditionally prescribed rules or dictates, without any consideration of the further consequences. The use of the term "dictates" implies including in this method the position according to which the ultimately valid moral imperatives are those referring to particular acts. Sidgwick himself, in Chapter 8 of Book I, alerts the reader to the different meanings he will assign to the term intuitionism, where those differences are due to the different generality of the intuitive beliefs recognized as ultimately valid.

The three methods analysed in *Methods* are not examined historically, as they are decision-making procedures that have effectively been proposed to govern everyday conduct, seeking to identify the changes that have come about over the centuries. Rather they are analysed insofar as, at least to the extent to which they are not mutually reconcilable, they represent alternatives from which human thought seems necessarily obliged to choose when

[26] Henry Sidgwick, *The Methods of Ethics*, p. 78.

it seeks to work out a complete synthesis of the practical maxims by striving to act in a perfectly consistent manner.

If, as it is often the case, the different common-sense methods applied in concrete circumstances provide mutually conflicting prescriptions, not all of them are acceptable:

[…] whereas the philosopher seeks unity of principle, and consistency of method at the risk of paradox, the unphilosophic man is apt to hold different principles at once, and to apply different methods in more or less confused combination […]. For if there are different views of the ultimate reasonableness of conduct, implicit in the thought of ordinary men, though not brought into clear relation to each other […] we cannot, of course, regard as valid reasonings that lead to conflicting conclusions; and I therefore assume as a fundamental postulate of Ethics, that so far as two methods conflict, one or other of them must be modified or rejected.[27]

Much of book IV of *Methods* is devoted to the attempt to harmonize and reduce to unity the different methods of ethics. However, I should like to point out that in the present article, in view of the objectives illustrated above, I shall not take into consideration the successful reconciliation between intuitionism and utilitarianism; instead, in view of the reconciliation between these two methods, I shall dwell on the problems raised by the attempt to seek a synthesis between utilitarianism and "rational egoism". Before directly addressing the problems linked to the relationship between these two methods, it is necessary to say something about the specific characteristics of "rational egoism".

3.2. *"Egoistic hedonism"*

Sidgwick devotes book two of *Methods* to the examination of "egoistic hedonism". He defines "egoism" as a method for determining the reasonable behaviour whereby each individual is supposed to adopt personal happiness as his own exclusive goal. Right from the outset, Sidgwick is aware of the innovative nature of the assumptions on which his investigation is based:

It may be doubted whether this ought to be included among received "methods of *Ethics*"; since there are strong grounds for holding that a sys-

[27] Henry Sidgwick, *The Methods of Ethics*, p. 6.

tem of morality, satisfactory to the moral consciousness of mankind in general, cannot be constructed on the basis of simple Egoism.[28]

He nevertheless deems it easy to dispose of this objection based on common-sense assertions that the principle has been widely accepted that it is reasonable for men to act in the way more likely to lead to their personal happiness:

Indeed, it is hardly going too far to say that common sense assumes that 'interested' actions, tending to promote the agent's happiness, are *prima facie* reasonable: and that the *onus probandi* lies with those who maintain that disinterested conduct, as such, is reasonable.[29]

According to the definition proposed by Sidgwick, it is necessary to define as egoistic the agent that, when faced with several possible lines of action, ascertains as accurately as possible the amount of pleasure and pain that is likely to result from each action and chooses the one which she believes will bring her the greatest happiness. The quantitative characterization of the rational goal of egoistic conduct deserves further clarification. The notion of the greatest possible happiness cannot be fully understood unless the meaning of "good on the whole" is clarified. A person's "good on the whole" is what she would desire and seek to achieve if she had fully understood all the consequences of all lines of conduct available to her. As Sidgwick perceptively points out, it is a terrible error to define a person's good simply as what would be desired if the outcomes of a given action could be predicted. It might always be possible that the choice of a particular object, while not emerging as an apparent good, that is, not different from what had been imagined, could on the whole be a bad choice owing to the concomitant aspects and long-term consequences. Sidgwick asserts that:

For it is not even sufficient to say that my Good on the whole is what I should actually desire and seek if all the consequences of seeking it could be foreknown and adequately realized by me in imagination at the time of making my choice. No doubt an equal regard for all the moments of our conscious experience – so far, at least, as the mere difference of their position in time is concerned – is an essential characteristic of rational conduct. But the mere fact, that a man does not afterwards feel for the consequences of an action aversion strong enough to cause him to regret it, cannot be accepted as a complete proof that he has acted for his 'good on the whole'. In-

[28] Henry Sidgwick, *The Methods of Ethics*, p.119.
[29] Henry Sidgwick, *The Methods of Ethics*, p. 120.

deed, we commonly reckon it among the worst consequences of some kinds of conduct that they alter men's tendencies to desire, and make them desire their lesser good more than their greater [...].[30]

Sidgwick claims that the principle prescribing that "one ought to aim at one's good on the whole"[31] must be considered as the self-evident intuition that underlies the "rational egoism" method. This principle is seen to be immediately self-evident when we consider individual goods of the person as similar parts of a quantitative or mathematical complex. In this perspective, the values of the individual goods will be assigned solely from the point of view of her maximum overall good, and the importance assigned to an individual good will be no greater than that which it has in the economy of her overall good. In other words, this principle states that a person must have an impartial interest for all parts of her conscious life. Of course, Sidgwick does not mean that a present good cannot reasonably be preferred to a future good on the strength of its greater certainty; he merely means to affirm that the mere difference of priority and posterity in time "is not a reasonable ground for having more regard to the consciousness of one moment that to that of another".[32]

Given Sidgwick's eudemonistic or hedonistic interpretation of the good, the principle of prudence may be expressed by stating that it is reasonable to forgo a present pleasure or present happiness in return for greater future pleasure or happiness or, more simply, that "a smaller present good is not to be preferred to a greater future good".[33]

From the foregoing the normative nature of the "egoistic hedonism" method emerges clearly: it consists in restricting a present desire in the wake of predictions of the more distant consequences deriving from such gratification.

The entire first chapter of book II is devoted to clarifying the notions of "interest" and "happiness", terms that in the author's opinion are too vague and ambiguous to be used in a scientific discussion on ethics. Sidgwick defines the notion of "greatest possible Happiness" as the "greatest attainable surplus of pleasure over pain"[34], where the term pleasure is used in its broader acceptance which includes all kinds of agreeable feelings: " the most refined and subtle intellectual and emotional gratifications, no less than the coarser and more definite sensual enjoyments"[35]. Acceptance of this quanti-

[30] Henry Sidgwick, *The Methods of Ethics*, p. 111.
[31] Henry Sidgwick, *The Methods of Ethics*, p. 381.
[32] Henry Sidgwick, *The Methods of Ethics*, p. 381.
[33] Henry Sidgwick, *The Methods of Ethics*, p. 381.
[34] Henry Sidgwick, *The Methods of Ethics*, p. 120.
[35] Henry Sidgwick, *The Methods of Ethics*, p. 127.

tative definition of the aim of egoism would imply that pleasures must be sought in proportion to their pleasantness, in such a way that the less pleasant state of consciousness cannot be preferred to the more pleasant state simply because the latter possesses some other qualities.

This conception of pleasure, which revisits Bentham's thesis of the complete homogeneity of pleasurable states of consciousness, completely contradicted the idea, defended by John Stuart Mill in his *Utilitarianism*[36], that it is possible to make a clear-cut distinction between qualitatively superior and qualitatively inferior pleasures. Sidgwick remarks:

This position, however, seems to many offensively paradoxical; and J. S. Mill in his development of Bentham's doctrine thought it desirable to abandon it and to take into account differences in quality among pleasures as well as differences in degree.[37]

According to Mill, differences in value between lower and higher pleasures were an "unquestionable fact"[38]. Sidgwick believed that the outlook defended by Mill could be accepted only if all the distinctions of quality could be resolved into considerations of quantity:

Now here we may observe, first, that it is quite consistent with the view quoted as Bentham's to describe some kinds of pleasure as inferior in quality to others, if by 'a pleasure' we mean (as is often meant) a whole state of consciousness which is only partly pleasurable; and still more if we take into view subsequent states. For many pleasures are not free from pain even while enjoyed; and many more have painful consequences. … and as the pain has to be set off as a drawback in valuing the pleasure, it is in accordance with strictly quantitative measurement of pleasure to call them inferior in kind.[39]

Sidgwick also believed that if non-hedonistic reasons for the preference were introduced into the egoistic calculation it would no longer be possible to consider egoism an autonomous method of ethics. Should it be admitted that the quality of the pleasures must be considered as something distinct from their quantity, and that it could even prevail over them, "egoistic hedonism" would no longer be clearly distinguishable from intuitionism.

---

[36] John S. Mill, *Utilitarianism* (1861), edited by Roger Crisp, Oxford, Oxford University Press, 1998, especially chap. 2.
[37] Henry Sidgwick, *The Methods of Ethics*, p.94.
[38] John S. Mill, *Utilitarianism*, p. 56.
[39] Henry Sidgwick, *The Methods of Ethics*, p. 94.

Before concluding this short reconstruction of the treatment of prudence as it appears in the *Methods*, I should like to examine what Sidgwick considered to be the difficulties implicit in the application of this method to cases of real conduct.

The fundamental assumption underpinning this method, which is implicit in the idea of considering a greater surplus of pleasure over pain as the ultimate aim of the conduct, is that all pleasures and all pains have a precise degree of positive or negative desirability which is knowable by the agents. Can it be assumed that in actual experience these degrees of desirability can be given with such precision? If this were false would it be a decisive objection to prudence?

Another assumption is that our pleasures can be increased and our pain decreased by means of forecasting and calculation. Nevertheless, it could be claimed that the practice of observation and hedonistic calculation inevitably tends to decrease our pleasures, at least the more important ones. It would thus seem problematic to try and attain our greatest happiness by attempting to pursue it scientifically.

Let us consider the latter objection first. Following Butler, Sidgwick affirms that it is possible to detect a difference between "extra-regarding" impulses and those whose object is our pleasure.[40] He also stresses that the greater part of our pleasure derives precisely from the satisfaction of those desires whose goals are different from pleasure itself.[41] In view of these premises it is easy to imagine what implicit danger lurks in the attempt to systematize conduct according to the principle of egoism: impulse towards our own pleasure could absorb the mind to such a degree as to become incompatible with the flow of those disinterested impulses towards particular objects, the existence of which is necessary in order to attain to a high degree that happiness toward which the principle of "egoistic hedonism" tends. This conclusion, which Sidgwick calls the "fundamental paradox of hedonism", must not be considered a decisive argument against this method:

I should not, however, infer from this that the pursuit of pleasure is necessarily self-defeating and futile; but merely that the principle of Egoistic Hedonism, when applied with a due knowledge of the laws of human nature, is practically self-limiting.[42]

---

[40] Henry Sidgwick, *The Methods of Ethics*, p. 44, see also p. 51.
[41] Henry Sidgwick, *The Methods of Ethics*, p. 44.
[42] Henry Sidgwick, *The Methods of Ethics*, p. 136.

In other words, according to Sidgwick, the only conclusion that can be drawn from the "paradox" is that the same method to achieve the end towards which egoism tends demands that to some extent we must place it outside our view and do not tend directly towards it. Once this danger has been clearly perceived it is no longer a cause of difficulty in the practical attainment of hedonism. As Sidgwick says:

For it is an experience only too common among men, in whatever pursuit they may be engaged, that they let the original object and goal of their efforts pass out of view, and come to regard the means to this end as ends in themselves: so that they at last even sacrifice the original end to the attainment of what is only secondarily and derivatively desirable. And if it be thus easy and common to forget the end in the means overmuch, there seems no reason why it should be difficult to do it to the extent that Rational Egoism prescribes [...].[43]

In Sidgwick's view, more serious objections may be raised concerning the possibility of performing precisely and reliably the methodical calculation of pleasure and pain required in order to adopt the method of egoism. In the first instance, if pleasure exists only insofar as it is felt, the fundamental assumption of egoism on the basis of which each pleasure has a quantitatively defined and measurable intensity must remain an a priori assumption that is not subject to any empirical verification. It is actually possible to assign a measure to a specific pleasure only when it is compared with other pleasurable sensations, but since this comparison can take place only in the imagination, it can only be hypothetically affirmed that, should it be possible for certain sensations to be felt simultaneously, it would be seen that one is more desirable than another in a definite proportion.

Second, even if it is taken for granted that each of our pleasures and pains can be measured precisely, the problem remains of whether we are in a position to know these quantities exactly. Indeed, even assuming we have an extraordinary predictive imagination, we would have to assume that during the measurement various different conditions were satisfied: 1) the mind would have to be in a perfectly neutral state in order to imagine all types of pleasure without bias for or against some specific sensation; 2) our capacity to enjoy certain specific pleasures must not change over time; 3) the assessment of the hedonic value of a past sensation must not be subject to error; 4) when we make use of the experience of others there must not be any difference between their sensitivity to the different types of pleasure and ours.

[43] Henry Sidgwick, *The Methods of Ethics*, p. 137.

3..3. *The dualism of practical reason*

Sidgwick believed that the numerous critiques that may be made to egoism do not make up a sufficiently strong argument to refute this method. Despite the difficulties involved, people are able to calculate their own pleasures accurately enough to satisfy the needs of their own lives. In fact, the strength of the normative reasons provided by egoism is never challenged by Sidgwick and it is precisely their universally binding nature that determines the failure of the foundational objectives of the *Methods*. In order to illustrate this point it must be borne in mind that Sidgwick, starting from the realization of the failure of "Mill's test" in favour of utilitarianism, comes to the conclusion that it is necessary to follow a method that is the opposite of the inductive one. One of the principal themes developed by Sidgwick in his *Methods of Ethics* is the demonstration that the grasp of self-evident first principles is essential for the rational foundation of utilitarianism. The construction of the utilitarian principle requires explicit recourse to two self-evident axioms. These are necessary to account for the universalistic dimension of which its specific nature is composed. The term "universalistic hedonism", which Sidgwick frequently uses as a perfect synonym of "utilitarianism", has the precise function of underlining this characteristic of universality.

The two axioms are those referring to the "principle of reciprocity" and to the relationship between the part and the whole. The former of the two states that "whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances"[44]. In other words, Sigdwick's idea is that unless there are significant differences among the agents or in the circumstances of actions, the same conduct is both morally valid and universally binding. The second axiom is represented by a universalization of the principle of the "egoistic hedonism" examined in the previous section, in which the relationship between the part and the whole is applied "from the point of view of the Universe"[45]. In short, as the egoist will consider her individual goods from the point of view of her maximum overall good, so the universalist hedonist will view her own good and that of others "from the point of view of the universe", from which the good of the single individual is important only insofar as it con-

[44] Henry Sidgwick, *The Methods of Ethics*, p. 379.
[45] Henry Sidgwick, *The Methods of Ethics*, p. 382.

tributes to the overall good produced in the universe. As Francesco Fagiani emphasized, in this view, "the overall goods of individuals appear as parts of a whole [...] to which they are subordinate and in which, all contributions being equal, the identity of the individual source from which the increase in the universal good comes is in no way significant".[46]

If, accepting Sidgwick's proposal, we identify the good with the non moral value of "pleasure" or "happiness", and if we accept the two self-evident axioms, utilitarianism is fully founded.

However, as Sidgwick himself points out quite "dramatically", the second of the two axioms is actually made up of two principles, the second of which may be rationally rejected even if the first is accepted. The first principle by itself provides the foundation of the "rational egoism" theory; only the acceptance of the second principle, that is, the consideration of one's own overall good as a part of the overall good of the universe, allows the egoistic dimension of ethics to be transcended by the universalistic one.

Sidgwick concludes his *Methods of Ethics* by acknowledging the fact that no rational argument exists that is capable of convincing those who have accepted egoism to accept the utilitarian prescription.[47]

Much of the contemporary discussion aimed at founding utilitarianism may be viewed as an attempt to come up with arguments that would allow, within the second axiom, to bridge the gap between the first and the second principle. Those who insist on the "separateness of persons" can only reject the second principle of the second axiom. Those who intend to develop Sidgwick's project further would have to propose a radical reappraisal of the notion of person by defending a conception of personal identity that is much less compact than the traditional one. However, once we decide to follow this path we will be forced to reconsider the categorical distinction between prudence and altruism. The philosophical reflections of Derek Parfit will be decisive for this new direction.

4. *Derek Parfit's "Self-interest Theory"*

Parfit describes his Self-interest or S Theory as a theory of individual rationality in which each individual is assigned the substantive objective of pursuing those outcomes that, given her set of desires, would allow her life

[46] Francesco Fagiani, *L'utilitarismo classico. Bentham, Mill, Sidgwick*, Napoli, Liguori, 1999, p. 53.

[47] In the final chapter of his *Methods*, Sidgwick affirms that the only possible way would be to postulate the existence of a utilitarian God who realizes the harmony between utilitarianism and prudence.

to unfold in the best possible way. In order to appreciate the peculiarity of this theory, it might be useful to imagine we could know all the desires of all persons – past, present and future. Moreover, each desire indicates both the person that has the desire and the time of her life in which it occurs (now, yesterday morning, in twenty years' time). In view of the enormous quantity of information involved, what would the rational course of action be? In other words, what desires should we take into greater account in deciding what to do?

Theories of rationality have suggested different answers to these questions. For instance, they may disagree as to whether it is rational to consider only our desires, or whether our future desires are to have the same weight as the present ones. The "Self-interest Theory" assigns significance only to the agent's desires and deems that those of the others can only indirectly influence the deliberative process that culminates in action. To use the technical jargon used in *Reasons and Persons*, this theory is agent-relative (it assigns to each individual a different substantive aim). Each of one's own desires directly provides the agent with a reason for acting and at any given moment the best rational action is dependent on the balancing of the relative weights of each of the reasons generated by those desires.

In the wake of Sidgwick's view of prudence, Parfit affirms that the force of these reasons is dependent exclusively on the intensity of the corresponding desires and thus the time at which they are perceived has no influence: future desires, according to S theory, must in themselves have exactly the same weight as we assign to our present desires. In Parfit's words, S is a temporally-neutral theory. Future events will be of less significance only if they are less likely to occur, but this does not mean that they are assigned less weight solely because, if they do take place, it will be later in time.

Parfit believes that it is possible to conceive of three equally plausible versions of this theory which differently interpret the meaning of best outcome. According to the "Hedonistic Theory", for each individual the best outcome is the one that ensures the greatest happiness. The various versions of this theory put forward different conceptions of happiness and of the ways of measuring it. In accordance with the "Desire-Fulfilment Theory", what is better for each individual is that which satisfies her desires throughout his life. On the basis of the "Objective List Theory", some things are good for us even if we do not desire them and bad for us even if we do not fear them. Different forms of this version exist according to what we consider to be good or bad.

These three theories coincide to a certain extent: they all agree in including happiness and pleasure among the things that enhance our lives and unhappiness and pain among those that worsen it. Without constraining him to choose among the three versions, this fact allows Parfit enormously to

simplify his treatment of "Self-interest Theory" by permitting him to discuss the "Hedonistic Theory" exclusively.

4.1. *How the "Self-interest Theory" may be self-defeating*

Parfit believes that numerous arguments may be constructed for the purpose of testing the plausibility of a moral theory or a theory of rationality. Among these, the simplest consists in demonstrating that a theory is self-defeating: this argument actually requires making no particular assumptions and in some cases is able to demonstrate that a theory fails on its own terms and must therefore be rejected.

Nevertheless, in the case of many theories, being self-defeating is not the same as demonstrating that those theories are unacceptable or must be rejected. In some cases this argument simply shows that a theory needs to be revised or extended, while in others it is unable even to demonstrate such a weak conclusion. In this section we will examine the outcome of this argument in the case of the "Self-interest Theory" according to Parfit's treatment. It will be highlighted how, although S is self-defeating, this in no way signifies a negative outcome for this theory.

In his article *Prudence, Morality and Prisoner's Dilemmas* [48] and at greater length in the first part of *Reasons and Persons*, Parfit identifies four ways in which a theory may be self-defeating: 1) a theory T is "indirectly self-defeating at the individual level" when it is true that, whenever someone attempts to achieve the objectives assigned to him by *T*, the latter are actually achieved less well on the whole; 2) a theory T is directly self-defeating at the individual level whenever it is certain that, if a person successfully follows *T* (that is, he succeeds in performing the act that, among those available to him, he more successfully achieves the objectives assigned to him by *T*), by this very fact he will act in such a way that the objectives assigned to him by *T* are achieved less well than if it had not followed *T* successfully; 3) a theory T is directly self-defeating at the collective level whenever it is certain that, if we all follow T successfully, for this very reason we will act in such a way that the objectives assigned to each one by T will be achieved less well than if none of us had successfully followed *T;* 4) a theory T is indirectly self-defeating at the collective level whenever it is true that, in the case that several persons follow the objectives proposed by T, those objectives are achieved less well.

[48] Derek Parfit, *Prudence, Morality and Prisoner's Dilemma*, "Proceedings of the British Academy" 65 (1979), pp. 539-64.

Since the "Self-interest Theory" is not a code of collective conduct, but a theory of individual rationality, the fact that it is indirectly or directly self-defeating at the collective level cannot be considered an objection to it. Parfit thinks it is easy to demonstrate that the "Self-interest Theory" is indirectly self-defeating at the individual level: for most people it is true that even if they never choose the line of action leading to a worse outcome, it would certainly be worse to be inclined to pursue one's own interest exclusively; it might be better to adopt another attitude.

It is worth emphasizing that the attitude responsible for the objection should not be interpreted as a set of self-interested motives always encouraging purely egoistic actions. Parfit, like Butler and Sidgwick before him, stresses that it is possible to pursue one's own personal interest by means of actions performed under the influence of altruistic motives or motives that are not directly self-interested:

Suppose that I love my family and friends. On all of the theories people affects what is in my interests. Much of my happiness comes from knowing about, and helping to cause, the happiness of those I love [...]. Suppose that I know that, if I help you, this will be best for me. I may help you because I love you, not because I want to do what will be best for me.[49]

Taking these explanations into account, Parfit believes that the best way to describe what it means for persons to have the attitude to pursue their personal interest is to affirm that, although often acting in pursuit of other more specific desires, they never do what they believe is worse for them. If this is true, these persons will explain themselves more clearly not by saying they have a disposition to pursue their own interest but by saying instead that they have the disposition never to go against it.

Let us now describe how, for an individual who adopts this disposition, S may be indirectly self-defeating. This would happen whenever a person, without ever going against her own personal interest, suffered a worse outcome than if she had adopted some other disposition. Even when persons succeed in never doing what is worse for them, the fact of never being willing to sacrifice their own happiness could be worse. Changing their disposition could prove more advantageous for them.

The following is one of Parfit's better known examples. Kate is a writer. Her greatest desire is for her book to be successful. Since the quality of her book is so important for her, she loves her work and her life appears to smile at her. If her desire to write the book was weaker, her work would be boring and her life, on the whole, would be negatively affected. Nevertheless, Kate,

[49] Derek Parfit, *Reasons and Persons*, pp. 5-6.

under the effect of her strongest desire is led to work so frantically and for such long hours that she ends up feeling exhausted and sometimes very depressed. As she is aware of this state of affairs, she is convinced that by working less frantically her book might be less successful but she would be happier, thus avoiding these periods of severe depression. If she accepts the "Self-interest Theory", thereby acquiring the disposition not to go against her own interest, Kate will come round to the idea that she should not overwork as by so doing she would do herself harm. This is an obvious case in which S would be self-defeating. Indeed Kate would always be able to avoid working at such a frantic rate only by tempering the intensity of her desire. This would represent an even worse outcome in terms of personal interest, since in this case work would be more boring for her and her life would be negatively affected. In Kate's case it is therefore obvious that never sacrificing one's own egoistic disposition can make things worse.

In this example the "Self-interest Theory" is self-defeating in its hedonistic version. If we were to accept the "Desire-Fulfilment Theory", we could reject Kate's idea that overwork is the cause of her problem: by working so hard, even though she wears herself out and occasionally suffers from depression, she manages to improve her book's quality. In this way, she ensures that her greatest desire is more fully satisfied. According to this "Self-interest Theory" this is a more satisfactory outcome for her.

For those who do not accept the hedonistic version of S, Parfit invents a different case. Let us imagine being lost in the desert and chancing to meet someone who can lead us back home in exchange for a certain sum of money. Let us imagine that we are unable to pay immediately, and that we promise to reward our rescuer as soon as we get home. Lastly, let us assume that we are transparent, that is, that we cannot lie without being caught. Since it would be worse for us to have to pay the agreed reward, if we know we are never willing to go against our own interest, we will never keep our promise to pay. Since we are transparent, also our would-be rescuer is also aware of this, and abandons us in the desert. For us it would have been better to be trustworthy, that is, to have the disposition to keep our promise even when to do so would make matters worse.

In the two cases described by Parfit, if an individual has the disposition of never going against her own interest, she makes the outcome worse. Parfit claims that this is true for most persons, for most of their lives. The question is – does this mean that the S theory is intrinsically false? Is this a sufficiently strong argument to reject the "Self-interest Theory"?

4.2. *The "Self-interest Theory" is not intrinsically false*

The objection in question would be fatal to the "Self-interest Theory" if it prescribed that persons should adopt the disposition never to go against their own interest. However, this would be an unacceptable thesis.

Parfit's argument is constructed on three theses underpinning the "Self-interest Theory". S claims that "for each person, there is one supremely rational ultimate aim: that his life go, for him, as well as possible" (thesis S1). When applied to acts, S claims both that each of us has most reason to do whatever would be best for himself (thesis S2) and that is irrational for anyone to do what he believes will be worse for himself (thesis S3). From the above three propositions a fourth thesis may be derived concerning the rationality of dispositions, that is, the set of motives that the "Self-interest Theory" prescribes that each agent should adopt. The fourth thesis claims that each agent should try to have or seek to maintain the best possible motives in terms of self-interest, that is this set of dispositions about which it may be affirmed that there is no other one that is better for her to have (thesis S4).

It is sometimes very difficult to know whether a set of motives may be causally possible, or whether it is one of the best in terms of S. Parfit nevertheless claims that there are also many cases in which a person knows that it would be better for her if her motives were to undergo some change: for such persons it may be true, as has emerged in the two preceding cases, that never to be willing to sacrifice one's self interest can lead to worse outcomes. Furthermore, in cases in which the person knows how to produce such changes, the thesis S3 implies that for these persons it would be irrational not to produce it, and that it would instead be rational to seek to have another disposition.

What these sets of motives actually are is partly a question of fact and the details of the response differ according to the different persons and the different circumstances of their lives: what we know in advance is only that it would be better for some persons if they were occasionally to go against their own interest and were willing to do what is worse for them. The limiting case is that in which for a person, under certain circumstances, it would be better to try and become completely irrational.[50]

Parfit claims that the "Self-interest Theory", although not intrinsically false, may nevertheless be refuted by means of an argument that challenges its very rationality. Before reconstructing this objection let us examine the conception of personal identity on which it is based.

---

[50] See the well-known case of *Schelling's Answer to Armed Robbery* in Derek Parfit, *Reasons and Persons*,pp. 12-13.

4.3. *Derek Parfit's personal identity theory*

In this section a brief outline is given of the central elements of the discussion of personal identity which makes up part three of *Reasons and Persons*. The essential arguments of Parfit's conception largely follow in the wake of the theses illustrated in numerous previous articles, and in particular those of his well-known article *Personal Identity*(1971).[51]

The two polemical objectives in that article, the thesis that personal identity is perfectly determined (the questions bearing on the identity of persons allow of only "yes or no" answers) and that according to which "what counts" when survival is at stake is personal identity itself, actually represent the central focus of the comprehensive discussion in *Reasons and Persons*.

Parfit claims that our view of the nature of persons and their continuing existence over time can be schematically presented as two theses: 1) persons are individual and ontologically non -reducible facts, whose continuing existence over time does not depend on (that is, it is not made up of) the existence of empirical, physical or psychological facts. From this it may be inferred as a corollary that the existence of the same person in two different times is a fact that is always perfectly determinable. 2) The continuing existence of these individual entities, that is, their numerical identity, is "what counts" when we are considering questions involving our survival. Numerical identity is the only thing that can justify the special interest we have in our existence and our future well-being.

Using a surprisingly large number of procedures, Parfit endeavours to demonstrate that what we are inclined to believe is not what we should believe because common sense has "a false view of the nature of personal identity"[52]. As an heir to that antisubstantialist tradition that had its first defender in Locke and in Hume its strenuous supporter, Parfit is defending a reductionistic or complex theory of personal identity which aims at reducing any discourse on the nature of persons to a description of the relations among classes of mental states that can be described "impersonally", thereby eliminating all forms of reference to subjectivity, to the point of view of the first person. He consequently puts forward a criterion of personal identity according to which our continuing existence consists in the recurrence of a relation of psychological connectedness and/or continuity among states of consciousness ("Relation R").

---

[51] Derek Parfit, *Personal Identity*, "Philosophical Review" 53 (1971), pp. 3-27.

[52] See Derek Parfit, *Lewis, Perry and the Matters*, in A.O. Rorty, *The Identities of Persons*, Berkeley, University of California Press, 1976, pp. 91-107. See also Derek Parfit, *Reasons and Persons*, chap. 10.

As it will be attempted to explain in the following sections, two highly innovative theses are implied in Parfit's conception. In the first place, the fact that the psychological connectedness is a relation that allows of variations in intensity means that it is also possible for cases in which our identity is indeterminate to occur. In the second place, if this thesis is accepted, it must be assumed that it is the relation of continuity and psychological connection between my present states and the future ones rather than personal identity *per se* what justifies the special interest in our future well-being.

Parfit's proposal has been interpreted as one of the most radical attempts ever made to eliminate the subject-person from the basic elements of the world. This contributed to making Parfit's reflections an essential point of reference both for those participating in the analytical debate who are interested in the general image of the person and for those involved in the discussion on the criteria of personal identity. It must be stated from the outset, however, that these two lines of reflection concerning Parfit, from our point of view, take on a significant, albeit limited, role. This is because our main interest lies not so much in the discussion of the nature of the person or in the way in which an answer to this question accounts for the thousand and one puzzles of personal identity, but rather in the consequences that Parfit's theory of personal identity has on the classical theory of prudence.

Parfit actually claims that close relations exist among the nature of persons and their identity over time and our reasons for acting. Once our shared opinions concerning personal identity have been changed we must consequently modify some of our beliefs concerning what we have most reason to do: we must reappraise out beliefs concerning rationality.

4.4. *Locke's legacy: the psychological criterion of personal identity*

The contemporary debate on personal identity is often characterized as referring to the principles which allow us to establish, for instance, that the person appearing before us is the same as the one we previously knew, where the principles sought must not be understood as mere pragmatic criteria (as, for instance, when the identity of a subject is established using his fingerprints), but refer to the justification of our identification procedures. As emphasized by Harold W. Noonan, it is possible in this connection to speak of the "logically necessary and sufficient conditions for which a person identified at a given moment is the same person as that identified at another"[53]. Nevertheless, it should be stressed that in this kind of investigation it is im-

[53] Harold Noonan, *Personal Identity*, London, Routledge,1989, p. 2.

portant to explicitly state the link between the search for such criteria and the more general, but no less exacting, question referring to the nature of the person. As Derek Parfit correctly points out we are confronted with two closely related issues: 1) What is a person's nature? 2) What is it that makes a person at two different moments one and the same person, or more precisely what is it that necessarily implies the continuing existence of each person over time?

Parfit claims that an answer to the second question is at least in part an answer to the first: the necessary characteristics of our continuing existence over time actually depend on our nature. In our examination of Parfit's position, for the sake of the explanation we shall not follow the order dictated by the logical priority of these questions but will deal with the nature of the persons after having answered the question of what is implied by their continuing existence over time.

Parfit defends a particularly sophisticated version of what in the contemporary debate is commonly defined as a psychological criterion. In very general terms, this conception states that personal identity implies the continuity of memory. This idea seems *prima facie* plausible because, it is claimed, it is precisely memory which makes most people aware of their own continuing existence.

The origins of this conception may be traced back to John Locke who, in Chap. XXVII of Book II of his *Essay concerning Human Understanding,* in several pithy pages, addresses what some believe may be considered the first comprehensive discussion concerning the criteria of personal identity which allow one to speak of a unitary subject that is continuous over time. For Locke the only fact that counts is the existence of direct memory connections, that is, memories of past experiences.[54] Parfit partly modified this Lockean conception. First of all, he considers that should no memory connections exist between two persons, let us say, between X today and Y twenty years ago, a continuity of memory may subsist just the same. This would be the case when a chain of linked memories exists between X and Y. This is a fairly frequent occurrence for the majority of people: every day they have memories of experiences they had the day before. It thus seems plausible to imagine that one of the conditions to be able to affirm that two persons at different times are the same person is that continuity of memory exists between them. Secondly, Parfit claims that the Lockean conception

[54] John Locke, *An Essay Concerning Human Understanding* (1689), edited with an introduction by Peter H. Nidditch, Oxford, Clarendon Press, 1975. For a detailed discussion on the influence of Locke's seminal ideas on the contemporary debate on the self, see Raymond Martin & John Barresi (ed.), *The Rise and Fall of Soul and Self: An Intellectual History of Personal Identity*, New York, Columbia University Press, 2006.

would however have to be corrected so as to take into account other psychological facts. As well as memories there are also other forms of direct psychological connection that necessarily have some weight in a personal identity criterion, such as desires, beliefs that are conserved over time, the connection linking an intention to the subsequent action in which it is implemented, salient features of a character, etc. Parfit terms all these kinds of direct psychological links "psychological connections".

Once Parfit modified the Lockean position along these lines, he places at the centre of his psychological criterion the relation of psychological continuity. Parfit defines this fundamental relation as the occurrence of chains of strong psychological connections. Here *strong* is meant to signify the existence of connections that are acceptable on average; for instance, an adult person might be called upon each day to recall at least 50% of the previous day's experiences, and so on for all the other types of psychological connections mentioned above.

Psychological continuity, unlike psychological connection, is a transitive relation and may therefore represent the personal identity criterion over time. This enables us to formulate the "psychological criterion" of Parfit's personal identity: (1) "psychological continuity" exists only when there are linked chains of strong connections. X today is the same person as Y in a previous moment only if (2) X is in psychological continuity with Y. (3) personal identity over time consists precisely in the occurrence of facts like (2).

4.5. *The "Reductionist" conception of personal identity*

According to Parfit's psychological criterion, personal identity over time merely implies different types of psychological continuity. Parfit affirms that this conception may be considered as a "Reductionist" theory of personal identity. In the latter the fact of the identity of a person over time, that is, her continuing existence, is deemed to consist solely in the occurrence of simpler, i.e. psychological, facts. These facts may be described in an impersonal way, that is without explicitly affirming that these were the experiences of a specific person. In other words, it is possible to describe all the psychological facts characterizing the mental life of a person in purely objective terms, in the third person, thereby eliminating all references to a first-person point of view.

Opposed to this theory are the "Non-Reductionist" conceptions of personal identity. In their stronger version they affirm that personal identity over time does not consist solely in physical and/or psychological continuity: it is an additional fact, distinct from the latter. It consists in the existence of a spiritual substance, a simple purely mental entity that accounts

for both the unity of consciousness in the various moments in time and the unity of a life as a whole.

In the "Reductionist" conception, persons are merely sets of experiences made up of relations of direct "psychological continuity" or by weaker forms of connection. In accordance with Parfit's well-known metaphor, they resemble clubs: entities that exist in a certain sense, but which are not included among the substantial elements of the world, as entities characterized by being centres of experience, but which are completely exhausted in the individuals that constitute them.

If we are seeking an example of the ontological depotentiation of the subject, suffice it to examine Parfit's description of dying, which seems to resemble more closely the break-up of a meeting than the irreparable loss of something. For Parfit: "Instead of saying, 'I shall be dead', I should say, 'There will be no future experiences that will be related, in certain ways, to these present experiences'"[55].

4.6. *"Reductionist" thesis: "what matters" is not personal identity*

Within the framework of neurobiological and neuropsychological research, the results of the clinical examinations performed on patients suffering from different types of disorder seem to cast doubts on the conventional image of ourselves as unitary and continuous entities. The conflict between the philosophical considerations triggered by several clinical cases and the common sense intuitions concerning the self is a topic that receives extensive treatment in part three of *Reasons and Persons*. According to Parfit's interpretation, the forms of "dissociation of consciousness" believed to take place in the case of so called "split brains", those in which the connections between the cerebral hemispheres have been surgically severed, provide strong arguments in favour of his reductionist conception. They are deemed to demonstrate that "what matters", that is, what justifies the special interest we feel in our future, is not personal identity but the relation of psychological continuity and/or connection (relation R).

Parfit imagines a radical case of ramification of the streams of consciousness in which the brain of an individual A is split and transplanted into that of his two brothers B and C. The latter will have a relation of complete psychological continuity with their donor. Both of them, after waking up after the operation, will believe they are the dead brother: they will have the impression of remembering having lived his life, will have his same desires and his same intentions.

---

[55] Derek Parfit, *Reasons and Persons*, p. 281.

In this imaginary case, "Relation R" ("psychological connectedness" and/or "psychological continuity") is configured as a bifurcation. However, personal identity cannot take on this form. The donor and his two brothers, thus constituted, cannot be the same person. Since the donor cannot be the same as two different persons, and since it would be arbitrary to say that only one of the two brothers is the same as the first one, the best way of describing this case is to say that neither person is A.

Unlike ordinary cases, in which personal identity is merely the occurrence of "Relation R" (indeed in practically all real cases R takes on the form of a one-to-one relation: that is, it exists between a person who currently exists and a future person), these are cases in which "psychological continuity" and "psychological connectedness" exist without identity.

The question might thus be asked of whether the lack of identity is really so important. Parfit's answer is negative: what really counts is the "Relation R" whatever its cause (in normal cases the persistence of the "Relation R" is guaranteed by the continuity of the central nervous system, which is indeed the natural cause).

Parfit illustrates this point by discussing an imaginary story. Let us imagine that a 'Star Trek' science-fiction-like device is used to scan my body and break it up into its component parts and then sends a signal to Mars by means of which a body identical to the original is recomposed. Subjectively speaking what happens is this: I press a button on Earth and immediately find myself on Mars. Assuming total psychological continuity I could say that the individual recomposed on Mars is identical to my self on Earth. Let us imagine, however, that a second copy of myself is sent to Saturn. For the reasons given above it is no longer possible to assert that I am identical to the individual sent to Mars. On the other hand, it is easy to believe that this lack of identity does not count for very much: after all, the situation is the same as before with the addition of a third person on Saturn. A few problems could conceivably arise because of the split (quarrels over possessions, love for the same wife, etc.) although, according to Parfit, the type of survival that I am guaranteed by psychological continuity in the ramified case is what I ought to assign value to.

If we accept Parfit's "psychological criterion", each ramification corresponds to the death of an individual A and the birth of two "Parfitian heirs", B and C, in his place, neither of whom is identical to A. But as Di Francesco rightly points out, this is a death in which no one actually dies: "the subject is actually not an added value vis-à-vis the continuity of the

experience and if the latter persists (albeit is multiplied), we have no reason to complain of the loss of anything real"[56].

4.7. *Refutation of the classical theory of prudence*

Now I shall present a comprehensive treatment of the argument by means of which Parfit believes it is possible to refute the "Self-interest Theory".

In S the "requirement of equal concern" is fundamental: a rational person ought to have equal concern for all parts of his own future. This means that each of us may attribute less importance to what may happen in the future only if this remoteness makes the event less probable. According to Parfit this thesis may be challenged on the basis of the reductionist conception. As emphasized in the preceding section, on the basis of Parfit's position, what fundamentally matters is psychological continuity and/or connectedness. In more than one point Parfit reiterates that both these relations play an important role in determining the special interest we attribute to our future. With these premises in mind, Parfit states a general thesis:

(C) My concern for my future may correspond to the degree of connectedness between me now and myself in the future. Connectedness is one of the two relations that give me reasons to be specially concerned about my own future. It can be rational to care less, when one of the grounds for caring will hold to a lesser degree. Since connectedness is nearly always weaker over longer periods, I can rationally care less about my further future.[57]

It should be noted that Parfit defends a discount rate referring not to time but to the attenuation of one of the two relations making up what has fundamental importance. Unlike the discount rate referring to time, this new discount rate is unlikely to be valid for the near future.

According to Parfit we must accept the thesis (C). Even if there are some exceptions, numerous relations must be judged less important when they occur with reduced levels of intensity: friendship, complicity, kinship, responsibility are but a few of the possible examples. Psychological connectedness must be considered in a like fashion. If we accept (C) we are rejecting the requirement of equal concern. This requirement is central to the "Self-interest Theory" and so we must reject this theory.

---

[56] Michele Di Francesco, *L'io e i suoi sé. Identità personale e scienza della mente*, Milano, Raffaello Cortina Editore, 1998, p. 195.

[57] Derek Parfit, *Reasons and Persons*, p. 313.

Parfit imagines that a defender of the theory S could retort that it is possible to modify the theory in question in such a way as to take his objection into account by incorporating a discount rate referring to psychological connectedness. According to this revised version, the dominant interest of a rational being ought to be that referring to her own future, although at that moment she might have less interest in those parts of her future with which she currently has a less close connection.

Parfit counters this response by arguing that this revised theory could not be considered a version of the "Self-interest Theory". Indeed the revised theory severs the fundamental link between S and the person's good on the whole. In the previous sections it has been shown how a central characteristic of the various formulations of the classical theory of prudence is that it is irrational for anyone not to do what she believes to be her good on the whole. In the revised theory, on the other hand, this thesis would have to be abandoned: if it is not irrational to be less concerned with certain parts of one's own future, it may not be irrational to do what is deemed worse in relation to one's own good on the whole.

As can be seen from the latter statement, the reply by the defender of S cannot be accepted and Parfit's objection is still decisive.

4.8. *The immorality of imprudence*

The outcome of Parfit's argument against the "Self-interest Theory" shows that it is no longer possible, as required by classical theory, to consider imprudent actions as irrational, since prudence cannot be equated with practical rationality. This means not having any more philosophical arguments to criticize imprudent actions. As Parfit affirms:

> If we believe that an imprudent act is not irrational, the charge 'imprudent' will cease, for many people, to be a criticism. It will become merely descriptive, in the way that, for many, 'unchaste' is merely a description.[58]

It therefore becomes necessary to seek a new theory that, by using a criterion other than rationality, will enable us to censure imprudent actions. Parfit suggests modifying our moral theory in such a way as to extend its application also to those actions that, in the past, were not the primary object of moral evaluation.

[58] Derek Parfit, *Reasons and Persons*, p. 318.

In this perspective, Parfit believes two different strategies may be pursued. He limits himself to describing them in very general terms, without actually choosing between them.

The first proposal consists of an appeal to consequentialism, and in particular to an agent-neutral principle of beneficence. If, in order to obtain lesser benefits in the present, an individual acts in such a way as to obtain greater hardship in her old age, she acts in a way that, when considered impartially, is the cause of worse consequences as it increases the quantity of suffering in the world, in accordance with this line of argument it may thus be affirmed that this individual acts in a morally deplorable way as her imprudence makes the outcome worse. As Parfit claims in *Reasons an Persons*, from the impartial perspective of consequentialism, "it is no excuse that the outcome will be worse only for me"[59].

Conversely, the second strategy consists in extending that part of moral theory that is "agent- relative". This involves our special obligations towards those with whom we have special relations: those with children, parents, patients, clients, are only few examples. It could be affirmed, Parfit goes on to say, that the relation between me in the present and me in the future sets up similar special obligations.

Parfit is aware that a revision of our moral conception in one of these two ways "would be, for many people, a large change in their conception of morality", since it seems to be a very deep common belief in our shared ethical thinking that "it cannot be a moral matter how one affects one's one future"[60]. However, if we accept a complex and reductionist account of the self this is the only strategy available. From this perspective we can no longer maintain that there is a categorical distinction between the way our actions affect our future self and the way they affect other selves. The only reasons that apply to these situations are the moral ones.

[59] Derek Parfit, *Reasons and Persons*, p. 319.
[60] Derek Parfit, *Reasons and Persons*, p. 319.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 109-136

Ordinary Moral Knowledge and Philosophical Ethics in Sidgwick and Kant

Massimo Reichlin
Facoltà di Filosofia
Università Vita-Salute San Raffaele
reichlin.massimo@hsr.it

ABSTRACT
Sidgwick considered Kant as one of his masters. However, he never devoted any systematic attention to Kant's ethical theory; moreover, in *The Methods of Ethics* he concluded that Kantian ethics is inadequate to guide moral life. I review Sidgwick's references to Kant in order to show that – along with basic differences – there are significant similarities in the main project of the two philosophers; and I suggest that, should Sidgwick have deepened his understanding of Kant, he might have realised that Kantian ethics offered a somewhat different way to accomplish the philosophical project he was interested in, that is, the systematisation of the morality of common sense through the establishment of certain moral axioms. I also suggest that Sidgwick's misunderstanding of the "formula of humanity" is at the heart of his final dismissal of Kant's ethics and that deepening his understanding of Kant might have led Sidgwick to revise his views on the rationality of egoism, thereby opening the possibility to solve the dualism of practical reason. Finally, I offer some speculations on the reasons why Sidgwick never attempted a thorough confrontation with Kant, suggesting that both his distaste for Kant's metaphysics and his Millian utilitarian bias deterred him from it.

1. *A Puzzling Relationship*

In the famous autobiographical note added to the sixth edition of *The Methods of Ethics*, Sidgwick declares Kant one of «my masters» (ME 7, p. xviii)[1] alongside with Mill; he describes his ethical project as a struggle «to assimilate Mill and Kant» (Ibid.), and says that his final reconciliation of utilitari-

[1] I will use the abbreviations ME 1 and ME 7 to refer to *The Methods of Ethics*, 1st edition (1874) and 7th edition (1907), both quoted from the Thoemmes reprint (Bristol 1996); OHE to refer to the *Outlines of the History of Ethics* (1886), quoted from the Hackett reprint (Indianapolis 1988); HSM to refer to *Henry Sidgwick: A Memoir,* by A. Sidgwick, E. M. Sidgwick (Macmillan, London 1906); and G to refer to Kant's *Grundlegung*, quoted from the English translation *Fundamental Principles of the Metaphysic of Morals* in I. Kant, *The Critique of Pure Reason, The Critique of Practical Reason and Other Ethical Treatise and The Critique of Judgment*, Encyclopaedia Britannica, Chicago 1952.

anism and intuitionism was reached in part through the realisation of the «perfect harmony» (ME 7, p. xx) between the Kantian principle and the utilitarian one. This late reconstruction is confirmed by a short paper written three years after the first publication of *The Methods*, where he already pointed out the centrality of the Kantian element in his ethical viewpoint: «I identify a modification of Kantism with the missing rational basis of the ethical utilitarianism of Bentham, as expounded by J. S. Mill»[2]. Given the emphasis with which he includes Kant among the main inspirers of his project, it comes as a surprise that Sidgwick never set out, in his long career as a philosopher and a university teacher, to address Kant's ethics with any detailed attention. Indeed, as noted by M. G. Singer, «his failure to come to terms adequately with Kant's ethics may be the most difficult thing to understand about his approach to ethics and the most serious deficiency in it»[3]. Such failure is particularly puzzling since: i) Sidgwick taught ethics constantly from the '60s to his death; ii) he devoted considerable attention to other influential moral philosophers, such as Bentham, Martineau, Grote, Green, Spencer and Stephen; iii) he published a number of essays on Kant's metaphysics and epistemology[4], and taught an entire course on the Critique of Pure Reason and the Prolegomena[5]. Why then did Sidgwick — apart from some passages in *The Methods* — never devote to Kant's ethics more than the few pages contained in the *Outlines of the History of Ethics*?[6]

The pages in the *Outlines* — it must be added — are indeed deeply inadequate, considering that Sidgwick could read German, that he was a very remarkable historian of philosophy, and that his Lectures on the Philosophy of Kant are profound and extensive. The brief summary contained in the *Out-*

---

[2] H. Sidgwick, *Mr Barratt on 'The Suppression of Egoism'* (1877), in *Essays on Ethics and Method*, ed. by M. G. Singer, Oxford University Press, Oxford 2000, pp. 27-28, here at p. 27. One earlier testimony is a 1866 letter in which he declares Kant's phraseology «quite a revelation to me», and, after having censured German Idealism as «a monstrous mistake», he concludes that «we must go back to Kant and begin again from him. Not that I feel prepared to call myself a Kantian, but I shall always look on him as one of my teachers» (HSM, p. 151).

[3] M. G. Singer, *A Note on the Content*, ibid., p. xlii.

[4] These are: *The So-Called Idealism of Kant*, «Mind» 1879; *Kant's Refutation of Idealism*, «Mind» 1880; *A Criticism of the Critical Philosophy*, part I and II, «Mind» 1883; and *Kant's View of Mathematical Premisses and Reasonings*, parts I and II, «Mind» 1883.

[5] *Lectures on the Philosophy of Kant and Other Philosophical Lectures* (1905), Thoemmes Press, Bristol 1996.

[6] The 1888 essay on *The Kantian Conception of Free Will* might here be added, though it in fact discusses the metaphysical underpinnings of Kant's conception, rather than his ethical theory qua talis.

lines, on the contrary, lends itself to criticism on several grounds: i) it never mentions the *Metaphysics of Morals* (though it implicitly refers to it in various passages)[7]; ii) it does not recall the central doctrine of the moral law as a fact of reason, stated in the second *Critique*; iii) it clearly misunderstands (OHE, pp. 274-5) the significance of the second formula of the categorical imperative (the "formula of humanity", more on which will be said later); iv) it never refers to such central ideas, in the Kantian perspective, as those of the autonomy of the will and of a universal kingdom of ends; v) it attributes to Kant the view that the belief in a moral government of the world is necessary to motivate moral action — a view Kant holds in the first *Critique*, from which Sidgwick quotes (OHE, p. 276), but repudiates in all his ethical treatises[8].

The lack of a direct confrontation with Kant's ethical thought clearly has to do with Sidgwick's classification of Kant as an intuitionist, as well as with his failure to acknowledge Kant's as a distinctive method of ethics[9]. This is again very surprising, since Kant's moral philosophy is doubtless very different from those of the British moralists, from Cudworth to Whewell, that are the paradigmatic exponents of the polemic target constructed by Sidgwick under the heading of "intuitionism" and discussed in Book III of *The Methods*. True, it could be argued that Sidgwick did show a certain awareness of the fact that Kant is not simply a member of the intuitional school; in fact, he writes that we can find «distinct traces of Kantian influence in Whewell and other writers of the intuitional school» (OHE, p. 271), and cautiously speaks of a particular affinity of Kant with Price (Ibid.; both emphases are added): these expressions may suggest that perhaps Sidgwick was not willing to rank Kant among the members of the intuitional school tout court. However he does seem to conflate Kantian ethics and intuitionism throughout

[7] E.g. OHE, pp. 274-5. *The Doctrine of Virtue* is instead quoted repeatedly in *The Methods*: see for example ME 7, III, 9, note 1; ME 7, III, 13, concluding Note and note 15.

[8] The same passage concerning the «glorious ideas of morality» as «objects of applause and admiration, but not springs of purpose and action» is quoted as representing the definitive Kantian position in the paper read to the Synthetic Society on February 25, 1898 (*On the Nature of the Evidence for Theism*, in HSM, pp. 600-608, at p. 605). The critical judgment on the treatment of Kant in OHE may be partly qualified by noting that Sidgwick's work is intended for English readers; for the modern period, in fact, it is «mainly confined to English ethics, and only deals with foreign ethical systems in a subordinate way, as sources of influence on English thought» (OHE, p. v). Not by chance, the last paragraph of the work, where the pages on Kant appear, bears the title "German influence on English ethics".

[9] As lamented by J. Rawls, *Kantian Constructivism in Moral Theory*, «Journal of Philosophy», 77, 1980, pp. 515-572, at p. 556.

The Methods, and explicitly declares Kant an intuitionist in at least one passage (ME 7, p. 366)[10].

Why Sidgwick never devoted more of his scholarly attention to Kant's moral philosophy is very difficult to investigate; I will be offering my tentative speculations in § 5. What may perhaps be more confidently said is that, should Sidgwick have deepened his understanding of Kant, he might have found that: i) Kant's ethics is not as inadequate to the task of giving «complete guidance» (ME 7, p. xix) to our moral life, as he finally came to believe; ii) Kant's project is much more similar to Sidgwick's than the latter thought, with particular reference to the relationship between ordinary moral knowledge and philosophical ethics. In fact, Kant's system offers a way of elevating the Morality of Common Sense into a system of philosophical ethics that is different both from the attempts of traditional "intuitional" moralists and from Sidgwick's problematic incorporation of that morality within the utilitarian system. I will not venture to say that, should Sidgwick have understood Kant more in depth, he would have become a Kantian; what the following pages are going to suggest is rather that he would have had to choose among two alternative ways in which to accomplish his own main project, that is, to provide a philosophical defence of the morality of common sense. And — for reasons that will emerge in due course — it is not wholly certain that he would have chosen the utilitarian one.

2. *The Relationship between Ordinary Moral Knowledge and Philosophical Ethics*

The project of *The Methods* is deliberately Socratic: through «impartial reflection on current opinion» (ME 7, p. xx), Sidgwick tries to bring consistency to the Morality of Common Sense of his era, just as Aristotle had done for the morality of fifth century B.C. Athens. Sidgwick clearly does not accept Common Sense as a definitive authority: he claims that «the aim of a philosopher, as such, [is] to do somewhat more than define and formulate the common opinion of mankind. His function is to tell men what they ought to think, rather than what they do think» (ME 7, p. 373)[11]. The aim of moral

[10] Another passage explicitly including Kant among the «intuitive moralists» occurred in ME 1, p. 303. The passage is modified in the following editions.

[11] The point is perhaps most clearly stated in a later essay: «though I have always been anxious to ascertain and disposed to respect the verdict of Common Sense in any ethical dispute, I cannot profess to regard it as final and indisputable: I cannot profess to hold that it is impossible for me ever to be right on an ethical point on which an overwhelming

philosophy is to correct and rationalise the morality of common sense in view of a more systematic construction: this can be effected by confronting it with genuine intuitions such as the Kantian principle of impartiality, the utilitarian principle of universal benevolence and the principle of rational egoism. The upshot of this procedure is well known: the alleged opposition between intuitionism and utilitarianism is in fact due to a misunderstanding, while a deeper opposition lingers between morality and rational egoism, i.e., the famous dualism of practical reason.

I think it important to stress the analogies that this Socratic project bears to the procedure followed by Kant, particularly in the Grundlegung — presumably a book very well known to Sidgwick. What Kant is here trying to do is in fact, first, to use the analytic method (see Preface) to extract, from what he calls the «common rational knowledge of morality» ("gemeine sittlichen Vernunfterkenntnis", note that for Kant this basic knowledge is already in itself rational), the very idea of duty, thus moving to a philosophical knowledge of morality (Section I); second, to search the principles of this philosophical morality, passing from «popular moral philosophy» to the «metaphysic of morals» (Section II); third, in a synthetic vein, to show that morality is not a «creation of the brain» but a reality, thus passing from the metaphysic of morals to the «critique of pure practical reason» (Section III). In other words, Kant is in fact assuming that morality exists, and that it is just like ordinary people conceive it; what he tries to do is to elucidate the concept of it that is implicit in ordinary moral knowledge, before trying to vindicate it rationally, by showing how pure reason can be practical.

The method employed by Kant is in fact different from Sidgwick's[12]. Kant does not provide a large review of the morality of common sense, in order to show both its strengths and its difficulties, as done by Sidgwick; he starts with what he considers the implicit understanding of common sense, relative to what is unconditionally good — i.e., the good will — and tries to bring out what is contained in this idea: that is, the idea of being subject to duty, which in turn means being subject to a law of reason that objectively and interpersonally constrains the satisfaction of individual inclinations and the pursuit of individual and collective happiness. This leads him to single out the categorical imperative, in the formula of universal law, as the fundamental principle of morality, not as a principle needed to systematise the plural-

majority is clearly opposed to me» (H. Sidgwick, *Some Fundamental Ethical Controversies* [1889], in *Essays on Ethics and Method*, pp. 35-46, here at p. 35).

[12] This contrast is very much emphasized by A. W. Wood, *Kantian Ethics*, Cambridge University Press, Cambridge 2007, pp. 43-65.

ity of moral imperatives acknowledged by ordinary moral consciousness, nor as one generated by a theorist's speculations, but as the principle that is ordinarily — though implicitly — used by common men; these, of course, do not conceive the principle in such an abstract and universal form as presented by Kant,

yet they always have it really before their eyes and use it as the standard of their decision. Here it would be easy to show how, with this compass in hand, men are well able to distinguish, in every case that occurs, what is good, what bad, conformably to duty or inconsistent with it, if, without in the least teaching them anything new, we only, like Socrates, direct their attention to the principle they themselves employ; and that, therefore, we do not need science and philosophy to know what we should do to be honest and good, yea, even wise and virtuous. Indeed we might well have conjectured beforehand that the knowledge of what every man is bound to do, and therefore also to know, would be within the reach of every man, even the commonest (G, pp. 260-261)[13].

In other words, starting from the idea, supposedly acknowledged by common sense, that the value of an action done from duty stems from its principle of willing and not from the object it pursues, Kant comes to the conclusion that the principle of «the moral knowledge of common human reason» (G, p. 260) — that is, the method used by ordinary men in reaching moral conclusions — is the one that tests moral maxims by asking whether they are the product of any inclination or are apt to become principles of a universal legislation.

On the contrary, Sidgwick embarks on a large review of the morality of common sense, in order to show that it does not provide a systematic construction, since many of its precepts are too vaguely stated and often at odds to one another. He then proceeds to extract from that large discussion three

[13] As is well known, Rousseau's influence was decisive for the development of Kant's ethical views on this point. In his *Notes* on his own copy of the *Observation on the Feeling of the Beautiful and the Sublime*, Kant already wrote (in 1765): «I am myself a researcher by inclination. I feel the whole thirst for knowledge and the curious unrest to get further on, or also the satisfaction in every acquisition. There was a time when I believed that this alone could make the honor of humanity and I despised the rabble that knows nothing. Rousseau set me right. This dazzling superiority vanishes, I learn to honor man and I would find myself more useless than the common labourer if I did not believe that this observation would impart to all else a value to restore the rights of mankind» (quoted in J. B. Schneewind, *The Invention of Autonomy. A History of Modern Moral Philosophy*, Cambridge University Press, Cambridge 1998, pp. 488-489).

or four immediately evident and more formal principles[14], that are genuinely axiomatic — that is, that are self-evident upon reflection for every rational individual — and that he finds partly in the works of past moralists and partly implicit in the ordinary way of dealing with moral questions. Armed with these principles, he goes on to show that they are able to provide the required systematisation of the morality of common sense, within the context of a revised utilitarian theory[15].

It is not the case that Kant, unlike Sidgwick, meant to withhold initial trust from the main normative principles that ordinary moral knowledge assumes to be true. On the contrary, in the Grundlegung, he seems to consider the fact that normative conclusions that we generally trust and assume to be true — e.g. that suicide is morally wrong, that we should not make promises with the intention not to keep them, and so on — can be derived from abstract formulations such as those of the various formulas of the categorical imperative as confirming that these formulas are in fact implied in the ordinary processes of moral thinking. Moreover, Kant's project in the later Metaphysics of Morals, which is his explicit attempt to construct a system of moral duties, is precisely to show the capacity of his philosophical system to vindicate most of the particular moral conclusions that were commonsense in his days and for his cultural and religious milieu. Kant's attitude towards the morality of common sense is in fact even more positive than Sidgwick's; he is however at least as sceptical as Sidgwick about the previous philosophers' attempt to provide a philosophical account of such ordinary knowledge. He therefore believes that the first philosophical task is to investigate the formal processes that are embedded in ordinary moral thinking, in order to establish its fundamental principle.

It might be observed that the difference between the two philosophers lies simply in the order in which the different steps are accomplished: Sidgwick just postpones the search for more formal principles after the review of common sense morality's material principles, but he nonetheless concurs with Kant in stressing the need for such principles in order to accomplish the systematisation that is philosophy's main task. This is not quite true, for it

---

[14] The question is notoriously controversial as to how many really self-evident principles Sidgwick is willing to accept: the figures range from three (H. Rashdall, *A Theory of the Good and the Right*, Clarendon Press, Oxford 1907, vol. I, p. 147) to eight (W. K. Frankena, *Henry Sidgwick*, in *Encyclopedia of Morals*, ed. by V. Ferm, Philosophical Library, New York 1956), with four perhaps being the most reasonable answer.

[15] Of course, he also shows the reasons why someone might not want to accept the principle of universal beneficence, thus confining himself to the narrower view of individual hedonism; this is what triggers the problem of the «dualism of practical reason».

seems to miss one important point: the fact that for Kant the formality of the fundamental principle is strictly connected with the formality of his conception of moral obligation. That is, while both philosophers clearly accept the idea that moral imperatives are dictates of rationality, Sidgwick assumes that they have to do with bringing about some good: in particular, the "ultimate good on the whole" must be identified with «what as a rational being I should desire and seek to realise, assuming myself to have an equal concern for all existence» (ME 7, p. 112). On this view, that choice is practically most reasonable which brings about the greatest good, however defined[16]. Kant, on the other hand, believes that, from the moral point of view, only a good will is unconditionally good, and it is good on account of its principle of willing, not on account of its object. Therefore, a deep difference between the two projects of vindicating philosophically the morality of common sense lies in the different conception of goodness that they assume as implicit in ordinary consciousness: on the one side, the idea of goodness as some state of affairs that can be produced — and that is eventually identified by Sidgwick with some pleasurable state of consciousness; on the other hand, the idea of goodness as good will, that is, the disposition to act only on maxims that may be conceived and willed as universal laws.

This basic difference should not prevent us from stressing the affinities between the two philosophical projects of founding a scientific ethics by giving philosophical systematisation and vindication to the morality of common sense. The Kantian way of proceeding is in fact doubly consonant to Sidgwick's mind: on the one hand, it shares its Socratic bent, by according serious philosophical relevance to the ordinary processes of moral knowledge; on the other hand, it clearly denies the sufficiency of ordinary moral knowledge for a genuine philosophical system of morality. In fact, while acknowledging that the common intellect may often surpass the philosopher in the practical domain, Kant declares that, lacking a precise philosophical determination of the principle of morality, it is difficult for ordinary wisdom to outdo the inclinations; that is, it is practically difficult to overcome the natural «disposition to argue against these strict laws of duty and to question their validity»,

[16] Both the axiom of prudence and that of beneficence are formulated by Sidgwick, by using the formal notion of "good"; they then receive "material" content through the demonstration that «Desirable Consciousness» is the only thing that can be considered as ultimate Good (ME 7, p. 397). But these formal readings would not escape Kant's objection that, by prioritising the good over the right, impure and heteronomous elements are introduced in the very concept of morality.

with a view, wherever possible, to «make them more accordant with our wishes and inclinations» (G, p. 261)[17].

On this account, it seems reasonable to say that Sidgwick might have found in Kant one alternative way of developing precisely what he wanted: the recognition of both the importance and insufficiency of ordinary moral knowledge, along with the philosophical effort to find a fundamental principle to systematize it[18]. The importance of the philosophical effort to bring systematic order into the ordinary moral knowledge of humanity is especially strong in both authors. As for Kant, his passion — or rather his obsession — for systematic philosophy, and for the critical foundation of the system of science, is too well known to be worth stressing. Let me just recall his observation that the innocence of practical wisdom is easily seduced; so that, «when practical reason cultivates itself, there insensibly arises in it a dialectic which forces it to seek aid in philosophy, just as happens to it in its theoretic use; and in this case, therefore, as well as in the other, it will find rest nowhere but in a thorough critical examination of our reason» (Ibid.).

As for Sidgwick, it is perhaps enough to quote a very strong passage from the I edition of *The Methods*, in which his quite rationalistic pretensions, as far as the foundation of morals are concerned, are very well voiced: «conduct appears to us irrational, or at least imperfectly rational, not only if the maxims upon which it is professedly based conflict with and contradict one another, but also if they cannot be bound together and firmly concatenated by means of some one fundamental principle. For practical reason does not seem to be thoroughly realised until a perfect order, harmony, and unity of system is introduced into all our actions»[19]. Doubtless, it is this epistemological ideal that renders Sidgwick's acknowledgment of the unsolvable dualism of practical reason so dramatic. He is notoriously emphatic about the uneasiness caused in him by the lack of a final foundation, and even considers a modification of his epistemology in order to close the gap between duty and happiness. After mentioning his previous willingness to accept a provisional postu-

[17] Schneewind appropriately stressed Rousseau's influence on this point as well; in fact, it is because human nature has been profoundly corrupted by its historical development, that feelings cannot by themselves reliably guide our action, and we need reason (*The Invention of Autonomy*, p. 504).

[18] Borrowing Schneewind's apt terms, we could say that Kant accepts the "dependence argument", but offers a "systematization argument" different from Sidgwick's (cf. *Sidgwick's Ethics and Victorian Moral Philosophy*, pp. 279-285 and 331-336).

[19] ME 1, p. 26 (the passage no longer appears after the second edition). On the strictly axiomatic character of Sidgwick's epistemic model, see J. Deigh, *Sidgwick's Epistemology*, «Utilitas», 19, 2007, pp. 435-446.

lation of immortality, in Kant's wake, while searching for the empirical evidence of an afterlife, he continues:

If I decide that this search is a failure, shall I finally and decisively make this postulate? Can I consistently with my whole view of truth and the methods of its attainment? And if I answer "no" to each of these questions, have I any ethical system at all? And if not, can I continue to be Professor and absorb myself in the mere erudition of the subject [...]. I am nearly forty-nine, and I do not find a taste for the old clothes of opinions growing on me (HSM, p. 467).

To decisively make the postulate would in fact be to accept the epistemology set out in the very last sentence of *The Methods*, according to which we would be justified in accepting as universally true propositions that are founded on our strong disposition to affirm them, together with their being indispensable to the systematic coherence of our beliefs. It is now a generally accepted view that Sidgwick never brought himself to make this postulate[20], even though he seems to waver significantly on this issue throughout his life: two years before his death, he in fact still seemed to consider such postulation a very serious possibility. The conclusion of his conference on theism is paradigmatic of his lingering doubts:

It seems to me, then, that if we are led to accept Theism as being, more than any other view of the Universe, consistent with, and calculated to impart a clear consistency to, he whole body of what we commonly agree to take for knowledge — including knowledge of right and wrong — we accept it on grounds analogous to those on which important scientific conclusions have been accepted; and that, even though we are unable to add the increase of certitude derivable from verified predictions, we may still attain a sufficient strength of reasoned conviction to justify us in calling our conclusion a "working philosophy" (HSM, pp. 607-608)[21].

[20] See for example the discussion in J. L. Mackie, *Sidgwick's Pessimism*, «Philosophical Quarterly», 105, 1976, pp. 317-327; repr. in B. Schultz (ed.), *Essays on Henry Sidgwick*, Cambridge University Press, Cambridge 1992, pp. 163-174.

[21] In the paper on *Authority, Scientific and Theological, presented to the Synthetic Society on February 24, 1899*, he returns on Kant's practical postulate again with a somewhat more sceptical attitude; in fact, «for most minds a belief recognised as assumed merely for practice is liable to decline to a belief of which there is an intellectual need, but a need that does not carry with it its own satisfaction: the satisfaction of the need has to be obtained, if at all, through some other line of thought» (HSM, pp. 608-615, here at p. 615).

In any case, it must be stressed that his aspirations to a philosophical foundation were such that his incapacity to solve this problem amounts for him to acknowledging a radical failure; as he confesses in 1887 note, «the recognised failure of my efforts to obtain evidence of immortality affects me not as a Man but as a Moralist» (HSM, pp. 471-472). In fact, while he does not feel anxious about the fact that, somehow or other, morality is going to get on, he sees clearly that, as a philosopher, his «special business is not to maintain morality somehow, but to establish it logically as a reasoned system; and I have declared and published that this cannot be done, if we are limited to merely mundane sanctions, owing to the inevitable divergence, in this imperfect world, between the individual's Duty and his Happiness» (HSM, p. 472)[22].

3. *Kant's search for the fundamental principle and Sidgwick's misguided critique*

Notwithstanding the affinity of the two authors' philosophical projects and some sparse similarities that will be noted in a while, Kant's moral deontology is doubtless very far from Sidgwick's utilitarianism «on an intuitional basis». I have already mentioned the fact that Kant assumes an anti-teleological notion of goodness at the very start of his philosophical inquiry on morality. Differences increase if we look at the development of Kant's fundamental principle, as spelled out in the second section of the Grundlegung. Here he distinguishes between hypothetical and categorical imperatives, and declares that moral imperatives command categorically. This implies that there are true answers to moral questions, that such answers can be found through rational reflection and that they are found by keeping such reflection "pure", that is, by discarding any empirical element, including of course the inclination towards certain objects.

Sidgwick does concur on part of this perspective. For one thing, he accepts a form of moral cognitivism, declaring, against the Humean view of reason shared by most part of the empiricist tradition, that «what ought to be is a possible object of knowledge» (ME 7, p. 33)[23]. And he intends this in the meaning of an objective rationalism, claiming that ethical judgments are

[22] This situation even led him to seriously wonder whether he had to resign his position as a teacher of ethics; see the 1888 letter in HSM, pp. 484-486.

[23] In the first edition he explicitly accepts the common view according to which «in saying that Reason apprehends moral distinctions, it would seem that no more is usually meant than that there is such a thing as moral truth and error; that two conflicting judgments as to what ought to be done cannot both be true and sound» (ME 1, p. 23).

«dictates» or «precepts» of reason, so that «what I judge ought to be must, unless I am in error, be similarly judged by all rational beings who judge truly of the matter» (Ibid.). Sidgwick's basic problem with the so called «dualism of practical reason» is in fact that, should we abandon the project of completely rationalising morality, cases of conflict between self-interest and duty would show practical reason as «divided against itself» and unable «to be a motive on either side»: the conflict would thus be adjudicated by the prevalence of one or the other group of non-rational impulses, and we should «lapse to the position which many utilitarians since Hume have avowedly held — that ultimate ends are determined by feeling, not by reason»[24].

This being so, it is also clear that Sidgwick does not accept only hypothetical imperatives, for he believes that reason also has a role in the determination of the ends, not only of the means. In fact, he distinguishes between "moral" and "prudential" judgments, meaning a distinction between «cognition or judgments of duty» (ME 7, p. 25) and «cognition or judgments as to what "is right" or "ought to be done" in view of the agent's private interest or happiness» (ME 7, p. 26). In the first edition, he even went so far as to refer this distinction to the one between an «authoritative, "categorically imperative" function of the Practical Reason» and «another in which its operation is more subordinate, prescribing not the end of the action but only the means to a given end. In this latter case the end is determined by desire or impulse of some kind, which may or may not be itself rational» (ME 1, p. 24). But the Kantian phraseology appears in some passages also in the last editions, for example where he says he wants to exhibit moral obligation as an «unconditional or categorical imperative» (ME 7, p. 35), and where he contrasts this categorically imperative interpretation of "ought" with the "ought" of the hypothetical imperative (ME 7, p. 37). Moreover, he declares that i) certain kind of actions «are commonly held to be right unconditionally, without regard to ulterior results» (Ibid.) and ii) that the same is true for the adoption of certain ends, such as the common good or general happiness. Lastly, there is also a very Kantian flavour in what today we would call the frank rationalistic internalism of Sidgwick's moral epistemology: to say that ethical judgments are dictates of reason, in fact, is for him also to say that «in rational beings as such this cognition gives an impulse or motive to action» (ME 7, p. 34), even though not always a predominant one.

Sidgwick's acceptance of the very notion of practical reason is indeed very rationalistic, anti-Humean, and generally foreign to the empiricist tradition; and Schneewind rightly suggested that there is a Kantian strain in Sidgwick's

[24] H. Sidgwick, *Some Fundamental Ethical Controversies*, p. 44.

notion of intuition «as the understanding a rational being has of the nature of his own activity as reasonable»[25]. This is in fact what seems to be implied in Sidgwick's notion that there are certain absolute practical principles, or axioms, «the truth of which, when they are explicitly stated, is manifest» (ME 7, p. 379): that, although there is no universal code of moral norms that are unconditionally valid for all human beings, there are certain formal principles that no rational being can ever deny, for they are, so to speak, consubstantial to the rational mind. However, it is also clear that Sidgwick does not accept Kant's specific idea of practical reason, that is, the idea that reason can be practical only by being pure, i.e. by putting aside all inclination and every other empirical element[26]. While it is clear that Sidgwick does not think of reason as a mere faculty of means, and of judgments as to what is right or ought to be done as mere instrumental judgments, it is also evident that, for him, the formal requirements of reason do not generate ends independently of any inclination; in fact, of the above mentioned absolute practical principles, he says that «they are of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case» (ME 7, p. 379). And he links the unconditional obligations he admits to the recognition of a universal end at which it is ultimately reasonable to aim, so that the obligation must concern acts mostly conducive to such end: in this way, he observes, «The obligation is not indeed "unconditional", but it does not depend on the existence of any non-rational desires or aversions» (ME 7, p. 35).

As for Kant, in the second section of the Grundlegung he strongly denies that happiness can be the source of the fundamental principle of morality, even though it is the only end that can be said to be pursued by all rational beings; the problem is that the imperatives of happiness command not necessarily, but only assertorially, in that they command something not for itself, but for something else, which we naturally will; moreover, such imperatives are consilia, rather than praecepta, for it is not possible to determine with certainty what will promote the happiness of a rational being at best. So Kant concludes that there is «but one categorical imperative, namely, this: Act only on that maxim whereby thou canst at the same time will that it should become a universal law» (G, p. 268).

---

[25] J. B. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy*, p. 420.

[26] See O. O'Neill, *Sidgwick on Practical Reason*, «Proceedings of the British Academy», 109, 2001, pp. 83-89. It could be argued, nonetheless, that Sidgwick's way of formulating the principle of universal benevolence in ME comes close to the purely formal indication of the others' happiness as an end that is at the same time a duty in The Metaphysics of Morals II, Intr., V. B.

This formula of the universal law is the one that Sidgwick constantly assumes as defining the Kantian position; according to him, the formula expresses the Golden Rule in a philosophically respectable form and gives general formulation to the idea of Justice as Impartiality (i.e. «that whatever is right for me must be right for all persons in similar circumstances», ME 7, p. xvii). Of this formula, however, he also says that it is «inadequate for the construction of a system of duties» (Ibid.), and that it does not «settle finally the subordination of Self-Interest to Duty» (Ibid.). In other passages — though with no direct reference to Kant — he adds that, strictly speaking, the effect of this principle «is merely to throw a definite onus probandi on the man who applies to another a treatment of which he would complain if applied to himself» (ME 7, p. 380); and he repeats the charge of insufficiency within the context of the administration of law, for «[the principle of impartiality] does not help us to decide what kind of rules should be thus impartially applied; though all admit the importance of excluding from government, and human conduct generally, all conscious partiality and 'respect of persons'» (Ibid.). Finally, he observes that the principle must be qualified by the belief that, in practice, the action whose maxim is being tested will not be widely imitated; otherwise, we should reject maxims such as the one to adopt celibacy, for, were it universally applied, it would determine the greatest of all crimes, i.e. the disappearance of the human kind. In short, the Kantian principle, for Sidgwick, «means no more that that an act, if right for any individual, must be right on general grounds, and therefore right for some class of persons; it therefore cannot prevent us from defining this class by the above-mentioned characteristic of believing that the act will remain an exceptional one» (ME 7, pp. 486-487).

While these contentions are fundamentally acceptable, a presentation of Kant's ethics, such as the one given by Sidgwick, centring only on this formula is in itself highly doubtful. It is true, of course, that Kant does say that «In forming our moral judgement of actions, it is better to proceed always on the strict method and start from the general formula of the categorical imperative: Act according to a maxim which can at the same time make itself a universal law» (G, p. 275). However, it cannot be forgotten that Kant does give at least two more formulas of the imperative; moreover, it is not at all clear that the "general formula" to which he refers in the passage just quoted must be identified with the formula of the universal law. In fact, it is clearly more sensible, and even more true to the letter of Kant's text, to interpret the sequence of the three formulations of the categorical imperative not just as the repetition of the same principle, but as a development of one central idea, progressively viewed from different perspectives. The ideas of humanity

as an end in itself, and of autonomy as the universally legislative will of every rational being, do in fact add much to the mere non-contradiction of the maxims, that is, to the purely formal condition set by the first formula. In view of this progressive development, there are serious reasons for the view that, when speaking of the "general formula", Kant is really intending the formula of autonomy, that is, the one that, in synthesizing the formula of the universal law and that of humanity, constitutes the most complete wording of the one fundamental principle[27].

Sidgwick never mentions the formula of autonomy, nor its variant centring on the kingdom of ends — not in *The Methods* nor in the *Outlines*; and shortly discusses the formula of humanity, both in the *Outlines* and in a long note in *The Methods*.

In the *Outlines*, Sidgwick introduces the formula of humanity after recalling Kant's thesis that ethics, unlike jurisprudence, is concerned with the realisation of internal freedom through the pursuit of rational ends, as opposed to the ends of natural inclination. Of Kant's statement that rational beings are ends in themselves, he notes that it is hardly a clear answer to the question asking what are the ends of reason. That statement might be interpreted as meaning that we should pursue the development of rationality, and therefore of morality, in every imperfectly rational being; but Sidgwick rightly dismisses this interpretation, since Kant clearly states that it would be a contradiction to promote the others' perfection: in fact, it is central to the attainment of intellectual and moral perfection that every man should autonomously pursue it. While we have a moral duty to cultivate ourselves, we cannot be morally bound to bring about others' perfection[28]. Having thus discarded perfection, Sidgwick sees no other way to interpret the formula of humanity than that according to which it commands to aim at the only other producible end of which Kant in fact speaks, that is, the happiness of others: therefore, everyone «is to help others towards the attainment of those purely subjective ends that are determined for each not by reason but by natural inclination» (OHE, p. 275).

In the *Note on Kant* concluding chapter xiii of Book III[29], Sidgwick links again the discussion of the formula of humanity to the attempt to establish the principle of beneficence. Here he notes that the derivation of this principle from the formula of universal law is not cogent, since we can clearly con-

[27] See A. W. Wood, *Kantian Ethics*, pp. 82-84.

[28] In ME 7, pp. 239-240, Sidgwick tries to show that this Kantian thesis is untenable.

[29] This note replaces the slightly larger discussion that appeared in ME 1, pp. 360-363 as § 4 of the same chapter XIII. The main passages here referred to are substantially identical.

ceive of a man who would prefer not to be aided by others to accepting obligations to aid, or who believed that a maxim of pure egoism is on the whole preferable, prudentially speaking. In this context, he expressly says that, for Kant, the fact that others are ends in themselves means that «we must recognise the duty of making their happiness our end» (ME 7, p. 389). Moreover, he reconstructs a different line of argument for the same principle, the one according to which, for Kant, since no particular object of inclination can be constituted as an absolute dictate of reason, only rational beings in themselves can be one such unconditional object or end. Then he goes on to criticise this argument, by noting that: i) to say that humanity is a self-subsistent end is perplexing, «because by an End we commonly mean something to be realised» (ME 7, p. 390); ii) there is a paralogism in saying that Men are ends in so far as they are rational and then deriving from this the duty to adopt as our ends their subjective, non rational ends: «It is hard to see why, if man as a rational being is an absolute end to other rational beings, they must therefore adopt his subjective aims as determined by his non-rational impulses» (Ibid.).

I will try to show in the next section that Sidgwick fails to understand the concept of humanity as an end in itself; this is not to be identified with a principle of beneficence according to which we are to make the subjective ends of others our own end, but with that self-subsistent end that grounds both our perfect and our imperfect duties towards ourselves and towards other (one of this last duties being the duty of beneficence). I will also argue that the failure to understand this key concept of Kant's ethics is at the heart of Sidgwick's substantial dismissal of the Kantian project.

4. *The interpretation of the formula of humanity as the key to Sidgwick's misunderstandings*

Two points must be stressed in Sidgwick's interpretation of the formula of humanity: the first is that Sidgwick clearly fails to grasp the meaning of the phrase "self-subsistent end". Of course, commonly an end is something to be realised; but Kant explicitly says that

since in the idea of a will that is absolutely good without being limited by any condition (of attaining this or that end) we must abstract wholly from every end to be effected (since this would make every will only relatively good), it follows that in this case the end must be conceived, not as an end to be effected, but as an independently existing end. Consequently it is conceived only negatively, i.e., as that which we must never act against and which,

therefore, must never be regarded merely as means, but must in every volition be esteemed as an end likewise (G, p. 276).

This conception of an end is perhaps uncommon in ordinary speech, but definitely not mysterious, and even standard in philosophical language: a self-subsistent end is simply a being already existing, for the sake of which something must be done, that is, a being that sets constraints on actions designed to produce any other end. It is not, therefore, an object of production but of respect; and humanity is such an end on account of the peculiarity of rational nature, that is, because of its capacity to set ends by herself, to have freedom and therefore moral agency. The idea of a self-subsistent end is simply the idea of an already existing source of the value of all ends, i.e., of a being whose unconditional worth grounds all conditional values. This notion of an end is explicitly worked out in Medieval thought, but has its roots in the Aristotelian idea of a final cause[30]; and it is surprising that Sidgwick, who knew very well Aristotle's work, did not grasp the meaning of this traditional idea.

The second point that must be emphasized is that, in commenting on the formula of humanity, Sidgwick fails to see that Kant is here after something which is central to his own enterprise. The formula, in fact, is not Kant's convoluted and perhaps inconsistent way of establishing the principle of rational benevolence, as Sidgwick seems to think; rather, it is his attempt to establish a much more fundamental principle that can be thought of as the basis of more specific normative principles, and, at the same time, as their limiting condition. Kant is actually giving substance, or matter, to his purely formal wording of the categorical imperative, as presented in the formula of the universal law. He seems indeed to agree with Sidgwick on the insufficiency of the formal principle, on its inadequacy for «complete guidance»; therefore, he complements it with a material principle, expressing the central value implicit in the formula of universality, that is, humanity as the capacity to set ends for oneself. Kant's principle of humanity therefore plays the same structural function as Sidgwick's universal benevolence, but has a much wider scope; it is the underlying principle at the basis of such diverse rules as the perfect duties not to commit suicide and not to make false promises, and

[30] The origin of Kant's phrase is traced by A. Donagan in the double notion of finis present both in Thomas Aquinas and John Duns Scotus (see *Human Ends and Human Actions: An Exploration in St. Thomas's Treatment*, Aquinas Lecture Series, Marquette University Press, Milwaukee 1985; repr. in Id., *Reflections on Philosophy and Religion*, Oxford University Press, Oxford 1999, pp. 81-97). Aristotle's main passage for the same idea occurs in the very famous chapter 7 of Book XII of the *Metaphysics* (1072a-1073a).

the imperfect duties to cultivate one's perfection and to pursue others' happiness; at the same time, it is the ground of the constraints on the imperfect duties. The road that leads Kant to this principle, as we saw, is different from the one followed by Sidgwick: while Sidgwick reaches the maxim of universal benevolence by showing that it is somehow implied in the ordinary rules and duties that he discusses in detail, Kant starts with the ordinary conception of good will and duty, showing that it implies both the idea of a universal legislation and of the autonomy of rational beings. Both philosophers, however, are looking for deeper principles, fundamental axioms or "intuitions", that can firmly concatenate practical maxims and adjudicate conflicts between them; and both understand such principles as expressing the fundamental nature of reason, and therefore as deeply embedded in our nature as rational beings.

What Sidgwick completely fails to see is therefore that the formula of humanity is the second step in the working out of the fundamental principle of Kant's ethics: this principle, in fact, is not — pace Sidgwick and most other commentators — the mere formal requirement that maxims should be universalisable[31]. Kant himself declares that maxims must have a matter or end; however, all material ends are relative and give rise only to hypothetical imperatives. Therefore, the source of a practical law, that is, of a categorical imperative, cannot be but in something whose existence has in itself an absolute worth, for «if all worth were conditioned and therefore contingent, then there would be no supreme practical principle of reason whatever» (G, p. 272). This something, which is rational nature in humans and other rational beings, is therefore an end, not in the sense of something to be produced, but in that of something to be respected that gives matter or content to the pure formality of a universal law: by rendering them moral beings, that is, capable of acting on the basis of the representation of a law, such matter or content in fact grounds the dignity of human beings, which, for Kant, cannot be replaced by something equivalent, for it is the source of all relative values, both of market values and of fancy values.

It is therefore correct to say that the formula of the universal law is insufficient for the construction of a whole system of duties; as Sidgwick notes, «all (or almost all) persons who act conscientiously could sincerely will the maxims on which they act to be universally adopted: while at the same time we continually find such persons in thoroughly conscientious disagreement as to what each ought to do in a given set of circumstances» (ME 7, p. 210). The

[31] This point is very well developed by A. Wood, *Kant's Ethical Thought*, Cambridge University Press, Cambridge 1999, pp. 182-190. Cf. Id., *Kantian Ethics*, pp. 85-95.

criterion of universalisability is a necessary but insufficient condition of morality, just because that formula is not the whole of Kant's fundamental principle. The principle is completely spelled out only when it is shown that all moral maxims, that is, all the maxims that can be justified from the moral point of view, have i) a form, i.e. universality; ii) a matter or end, i.e. rational nature as «the condition limiting all merely relative and arbitrary ends» (G, p. 275); and iii) «A complete characterization of all maxims by means of that formula, namely, that all maxims ought by their own legislation to harmonize with a possible kingdom of ends as with a kingdom of nature» (Ibid.). The third condition, which sums up the first two formulas through the idea of autonomy and of a kingdom of ends, is that every maxim should aim at a systematic connection of rational beings in a sort of kingdom; such a kingdom is defined by the fact that everyone is at the same time a legislator and subject to laws (i.e. universal objective principles) devised to treat everyone not as a means only but always at the same time as an end; this is the final and really complete formulation of Kant's fundamental principle — a formulation to which Sidgwick never makes reference throughout his work.

Understanding the role of the formula of humanity in the context of Kant's ethical system also helps to solve the specific problem posed by Sidgwick with reference to the rule of benevolence. In fact, the formula of humanity is in the first place a limitative condition on the acceptability of maxims; its main practical effect is to reject maxims that would allow treating rational beings as mere means for others' ends. Moreover, the formula has positive implications in suggesting the adoption of the two ends that, according to The Metaphysics of Morals, are at the same time duties: one's perfection and others' happiness, that is, the duties that in the Grundlegung appeared as examples of the imperfect duties towards oneself and towards others[32]. When we come to the positive part, however, we must not forget that rational beings are considered as ends in themselves qua rational. Therefore, while limiting oneself to withholding disrespectful actions would be to value rational nature too poorly, only in a negative fashion, adopting others' subjective ends is always constrained by the moral non-rejectability of such ends: when Kant says that «the ends of any subject which is an end in himself ought as far as possible to be my ends also» (G, p. 273), the possibility of

[32] It may plausibly be contended that the notion of the two ends that are at the same time duties, put forth in the *Metaphysics of Morals*, constitutes a real development with respect to the *Grundlegung*, where in fact this notion does not appear. However, the development may be interpreted as the working out in detail of the practical impact of the more abstract principle of the *Grundlegung*. On this issue see D. Tafani, *Il fine della volontà buona in Kant*, in L. Fonnesu (ed.), *Etica e mondo in Kant*, il Mulino, Bologna 2008, pp. 145-163.

which he is speaking is moral, not merely physical. In fact, in the later Metaphysics of Morals, he says that love is a maxim of benevolence, resulting in beneficence, this consisting in «the duty to make others' ends my own (provided only that these are not immoral)»[33]. In other words, respecting rational nature in any individual does include promoting those contingent ends that make up her life project and from which she can expect her happiness; but i) this is not the main meaning of such respect, and ii) this promotion is in any case constrained by the prior acceptability of those ends[34].

Sidgwick is therefore wrong when he points to the almost complete coincidence (ME 7, p. 385) between Kantian ethics and utilitarianism, based on the fact that «the only really ultimate end which he [i.e. Kant] lays down is the object of Rational Benevolence as commonly conceived–the happiness of other men» (ME 7, p. 386)[35]. Actually, not only does Kant add the other imperfect duty to cultivate one's intellectual and moral perfection; but Sidgwick also forgets that the duty of beneficence, as well as that of perfection, is limited by the negative part of the respect for rational nature as an end in itself. Kantian beneficence is therefore significantly unlike utilitarian one, for it is not the unlimited pursuit of others' subjective aims, as determined by their natural inclinations, but the pursuit of those subjective aims that pass the scrutiny of rational reflection, that is, that are not to be rejected on the basis of the two tests of universalisability and non-exploitation. The difference between Kantian and utilitarian beneficence was not grasped by Sidgwick, because he did not really understand the notion of a self-subsistent end, nor the role of the formula of humanity as a fundamental axiom or principle grounding those of perfection and benevolence, and not reducible to the last

[33] I. Kant, *The Metaphysics of Morals*, Cambridge University Press, Cambridge 1996, *Doctrine of the Elements of Ethics*, § 25, p. 199.

[34] It must be noted that Sidgwick's interpretation of the formula of humanity has shaped its understanding by many contemporary commentators: paradigmatically, it is adopted by R. M. Hare, *Could Kant Have Been A Utilitarian?*, in *Sorting Out Ethics*, Oxford University Press, Oxford 1997, pp. 147-165. In this same perspective, M. Nakano-Okuno recently went so far as to affirm not only that «this formula shares essentially the same claim as the Principle of Rational Benevolence», but also that it «encompasses the essential claim of Sidgwick's Principle of Rational Prudence», since it imposes to treat one's future ends as they were present (*Sidgwick and Kant: On the So-Called "Discrepancies" Between Utilitarian and Kantian Ethics*, in P. Bucolo, R. Crisp, B. Schultz [eds.], *Henry Sidgwick: Happiness and Religion*, Dipartimento di Scienze Umane, Università degli Studi di Catania, Catania 2007, pp. 260-333, here at p. 292 and p. 294).

[35] In a passing note of the first edition, he even committed himself to such an absurdity as to say that «in fact, as we have seen, [the utilitarian first principle] is the first principle of Kantism» (ME 1, p. 440).

one. Moreover, he could not realise the resources of the formula of humanity in adjudicating between conflicting grounds of duty: in fact, the formula clearly seems to justify a relative priority of perfect duties over imperfect ones, while leaving the last word to the exercise of judgment in the circumstances[36]. Finally, it is Sidgwick's failure to understand and to accept such notions as humanity as an end in itself, human dignity and moral autonomy that explains one of the most critical points of his ethical views, probably the one most often quoted in recent debate: his refusal to link moral reasonableness to the demands of publicity, and his consequent acceptance of utilitarianism as an esoteric morality reserved to the enlightened few[37]. While both philosophers emphasize the role of ordinary moral knowledge, Kant is in fact much more open-minded and 'progressive' in his recognition of the intellectual and moral competence of ordinary people; though he does not bring out all the consequences of his notion of morality as self-governance, his ideas of humanity and human dignity made him stand well over Sidgwick's still elitist morality. In the end, as noted by Schultz, the strongest difference between the two thinkers perhaps lies in the fact that it is hard «to find in Sidgwick's idea of a method of ethics an effectively Kantian endorsement of the plain person's capacity for moral self-direction»[38]. The reason of this difference probably lies in the different historical and cultural contexts of the two philosophers: while Kant, who wrote in the age of Enlightenment and in the context of the hopes generated by the American and French revolutions, had moderately optimistic views on history and the potentialities of human development, Sidgwick can be considered a sort of critic of the Enlightenment, and his elitist conclusion is the effect of the disbelief in any optimistic philosophy of history and in the reality of moral human progress.

---

[36] This seems to be implicit in the very tentative casuistry sketched by Kant in *The Metaphysics of Morals*. Also the *Lectures on Ethics* often testify to a somewhat more flexible attitude (even as far as truth-telling is concerned) on immediately practical matters than is generally thought.

[37] B. Schultz appropriately notes that «it is perhaps at this juncture that one can best appreciate how Sidgwick parted from the Kantian project» (*Henry Sidgwick: Eye of the Universe. An Intellectual Biography*, Cambridge University Press, New York 2004, p. 264). The centrality of the idea of publicity is particularly stressed, in a Kantian and anti-utilitarian vein, by J. Rawls, *Political Liberalism*, Columbia University Press, New York 1993, pp. 66-71. The charge that Sidgwick's esoterism amounts to "Government House" utilitarianism is notoriously put forward by B. Williams, *The Point of View of the Universe: Sidgwick and the Ambitions of Ethics*, in *Making Sense of Humanity and Other Philosophical Papers*, Cambridge University Press, Cambridge 1995, pp. 153-171.

[38] *Henry Sidgwick: Eye of the Universe*, pp. 267-268.

5. *Two speculations*

Sidgwick's avowed incomprehension of both the formula of humanity and the third formula (autonomy/kingdom of ends) is the key to understanding his final rejection of Kant's system, and his interpretation of Kant as an intuitionist[39]: this incomprehension explains his purely formalistic reading and his strategy to complement the formal principle of universalisability (justice, or impartiality) with the substantive principle of rational benevolence.

A further problem remains, as to why Sidgwick never tried to deepen his understanding of the whole Kantian project in ethics. On this, no more than speculations may be offered.

One tentative answer might be that Sidgwick was deeply convinced of the untenability of Kant's theoretical philosophy, and that this conviction deterred him from embarking in a serious study of Kant's ethics. There is in fact at least one passage in the Lectures on the first Critique in which Sidgwick stresses the uncomfortable consequences of Kant's epistemology for ethics. Speaking of the ideality of time in the transcendental aesthetics, he notes that this doctrine has the effect of rendering intellectual and moral progress mere appearances:

> Hence the conception of moral progress, on which the practical postulate of immortality — as we saw — is based, is a conception that represents no real fact of any soul's existence, but merely an appearance due to the imperfection of its faculty of cognition. But if moral progress is thus reduced to mere appearance, what becomes of the belief in the immortality of the soul which Kant (in the Critique of Practical Reason) bases on it? Indeed, in any case, if Time is merely a form of human sensibility, — due to an imperfection of man's nature which prevents him from knowing things as they are, — the postulate of immortality seems to become a postulate for the endless continuance of an imperfection. It does not seem that this can afford an inspiring hope for a truth-loving mind[40].

---

[39] It is curious that Sidgwick never puts any emphasis on the only passage lending true credibility to an intuitionist reading of Kant: the passage in the second *Critique* speaking of the moral law as a "fact of reason". To my knowledge, Sidgwick never mentions that passage in his works; most of his references to Kant are to the *Grundlegung*, and certainly the second *Critique* is never mentioned in *The Methods*, but for a note in passing at the beginning of the last chapter of ME 1, p. 439 — a note that was subsequently removed.

[40] *Lectures on the Philosophy of Kant*, p. 36.

And although he adds that, in his practical philosophy, Kant seems to defend a sort of noumenal freedom of the soul, according to which «the momentous choice between good and evil which every human soul makes is in reality not subject to the condition of time, so that any change that may appear in a man's character is illusory»[41], it is clear that in either way the metaphysical underpinnings of Kant's position were deeply uninviting for Sidgwick.

This explanation, however, is far from satisfying. For one thing, it is unclear why Sidgwick, as an historian of philosophy, should not have wished to deepen his understanding of Kant's moral doctrine, as he had done with the epistemological ones, even knowing that it was based on metaphysical grounds utterly unpalatable for him. Moreover, Sidgwick might well have done with the notions of humanity, autonomy and the kingdom of ends what he had done with the notion of universalisability: that is, to accept what he found useful in the normative principle, while discarding its theoretical underpinnings. This is in fact how he describes his attitude in the auto-biographical note: «What commended itself to me, in short, was Kant's ethical principle rather than its metaphysical basis» (ME 7, p. xvii).

Another possible explanation is that Sidgwick approached Kant's ethics with a serious bias deriving from his previous acceptance of Mill's utilitarianism, and that, although he subsequently tried to revise his first impression by rereading Kant, he never fully succeeded in developing an unbiased analysis. Owing to the prejudices he had inherited from Mill, he never got convinced of the necessity of a deeper understanding of Kant's ethical project, and this may explain the lack of a direct confrontation with it. In other words, Sidgwick started as a utilitarian and never ceased to be one; the difficulties he found in the theoretical frameworks of the masters of his school led him to reappraise the importance of common sense morality, as well as of authors such as Kant, Clarke and Butler: but he never ceased to think of utilitarianism, however revised, as the most satisfactory moral theory (at least, faute de mieux). So, he wanted to be a utilitarian — perhaps just as he wanted to be a Christian, though he carried out the latter endeavour less successfully,— and tried what he could in order to rescue utilitarianism from its defects; perhaps this attitude was also suggested by the Millian conviction that utilitarianism was the theory associated with moral and social progress, while all other theories were, in some way or other, conservative.

This explanation might be not particularly respectful of Sidgwick intellectual honesty; however, it has some textual evidence in its favour. Not only Sidgwick does start his auto-biographical note declaring that he adhered to

[41] *Ibid.*, p. 37.

Mill's utilitarianism from the start; he also adds that, in his subsequent search for a deeper foundation of ethics, he retained a general «attitude of discipleship to Mill» (ME 7, p. xvi); he says that «through all this search for principles I still adhered for practical purposes to the doctrine I had learnt from Mill» (ME 7, p. xviii); and he explicitly declares that the first time he read Kant, he read it «somewhat unintelligently, under the influence of Mill's view as to its grotesque failure» (ME 7, p. xvii). True, he also adds that he re-read it «more receptively» (Ibid.), discovering the importance of its fundamental principle. However, it is clear that even this second reading was in fact influenced by Mill: the idea of Kant's ethics as a merely formalistic system, and the neglect of the second and third formulas on which it is based, though commonsense in most philosophical literature, are central in Mill's reading; and the idea that the only end laid down by Kant through his imperative is rational benevolence is also near to the main tenet of Mill's interpretation. In the passage on the «grotesque failure» of Kant's ethics quoted by Sidgwick, Mill says that the only contradiction that the test of universalisability is able to detect is the one between certain immoral rules of conduct and the general desires of humanity; that is, the Kantian principle is meaningful, and able to justify duties of morality, only if interpreted in consequentialist terms, as rejecting the maxims on account of their consequences. Moreover, in the other passage of Utilitarianism in which Kant is mentioned, Mill says that the only meaningful sense of Kant's fundamental principle is that «we ought to shape our conduct by a rule which all rational beings might adopt with benefit to their collective interest»[42]; here again Mill interprets what Kant intends as an a priori constraint on moral maxims as a concern for the consequences of moral rules. Sidgwick's misunderstanding of Kant is similar: he likewise fails to appreciate the fruitfulness of the categorical imperative, in its three progressive formulations, and insists that it does not justify any moral rule per se, though it is the first step in the process of justifying the utilitarian principle of rational benevolence. In short, notwithstanding his testimony that he reread Kant's ethics «more receptively», Sidgwick seems to have been receptive only to the possibility of incorporating the Kantian idea of the universality of morality into the utilitarian system; that is, he complemented Mill's attempt to show the compatibility of Kant's

[42] J. S. Mill, *Utilitarianism* (1861), in *The Collected Works of John Stuart Mill*, Volume X - *Essays on Ethics, Religion, and Society*, edited by J. M. Robson, University of Toronto Press, Toronto 1985, pp. 205-260, here at p. 249. The passage on Kant's failing «almost grotesquely» is at p. 207.

ethics and utilitarianism, a compatibility that embodies a misunderstanding of Kant very similar to Mill's.

6. *Conclusive remarks*

My main contention has been that Sidgwick did not fully grasp the heart of Kant's moral project. This means that he did not have the opportunity to discover that Kant's ethics was much more congenial to his mature thought on ethics than he assumed, and also that, for certain aspects, it was much less conservative than most "dogmatic intuitionism"; in short, he did not have the opportunity to consider an alternative and plausible way to accomplish the philosophical task that he considered decisive. This far, we have not suggested that, should Sidgwick have understood Kant more deeply, he would have taken up the Kantian way, or that he could have been (some sort of) a Kantian. It is of course possible, and indeed likely, that Sidgwick's prior acceptance of a strictly teleological conception of ethics, according to which for an action to be morally appropriate is to maximally promote some good, would have prevented him from accepting both Kant's general theory and more specific ideas such as the conception of beneficence.

However, we can also consider some aspects that may have recommended to him the Kantian solution. The first point to consider is that the deontological conception prioritising the right over the good seems to be embedded in the morality of common sense in a way that the one-sided utilitarian insistence on consequences seems not. Of course, there are cases in which utilitarianism can be easily accorded with our considered judgments; but the central cases discussed by Sidgwick in Book III, those that are the traditional object of the utilitarian polemic, are not of this sort. Let us take the classical example of promises. Throughout his discussion, Sidgwick basically aims at showing that common sense is in many case uncertain as to the boundaries of the duty to keep promises. However, he clearly, though implicitly, acknowledges that the traditional casuistry — such as it had been revived and systematised by Whewell[43] — had defined several precise conditions for the treatment of hard cases, such as that a promise is binding «if the promiser has a clear belief as to the sense in which it was understood by the promisee, and if the latter is still in a position to grant release from it, but unwilling to do so, if it was not obtained by force or fraud, if it does not conflict with definite a priori obligations, if we do not believe that its fulfilment will be harmful to the

[43] W. Whewell, *Elements of Morality, Including Polity*, III edition, Parker, London 1854.

promisee, or will inflict a disproportionate sacrifice on the promiser, and if circumstances have materially changed since it was made» (ME 7, p. 311). Nowhere does this traditional treatment refer to the mere balance of good versus bad consequences in order to solve particular problems; on the contrary, it always keeps to the deontological intuition according to which principled, a priori solutions, not mere cost-benefit analyses, are needed also for hard cases. Here, as in other cases, Sidgwick's thesis that common sense lacks clear answers is perhaps right, but the same cannot be said of the systems of intuitionist philosophers such as Whewell, whose solutions Sidgwick simply omits to discuss[44]; his conclusion that precise duties can be defined, and conflicts of duties resolved, only by reference to the principle of utility, is therefore unwarranted. Moreover, to expect that the application of the principle of utility would bring to the treatment of promises much more definition and "scientific" precision than afforded by traditional casuistry is both to require from moral philosophy much more than it is legitimate (witness the same Aristotle that Sidgwick is imitating) and to overstate the utilitarian ability to predict specific consequences in particular cases. Finally, on the basis of his alleged aim to provide a philosophical foundation for the morality of common sense, Sidgwick himself ought have been sympathetic to the efforts of ethical theories that tried to treat hard cases and alleged conflicts of duties without giving up the deontological intuitions at the heart of ordinarily acknowledged duties: if not Kant's, at least the more philosophically refined formulations of the so-called "intuitionistic theory", such as Whewell's. But Kant's theory, correctly understood, offered precisely what Sidgwick was looking for: a philosophical principle, developed in the three stages spelled out in the Grundlegung, that can both systematise the rules of common sense morality and provide principled ways to adjudicate conflicts between them.

The second point to consider is the conflict between happiness and duty, that is, what Sidgwick styled the dualism of practical reason, and what he considered the most serious problem of ethics. Sidgwick's dissatisfaction with Kant's solution to this problem is well known; in a passage of the Memoir already mentioned, he recalls that, when writing *The Methods*, he was «inclined to hold with Kant that we must postulate the continued existence of the soul, in order to effect that harmony of Duty with Happiness which seemed to me

[44] On Sidgwick's unfairness in pointing to the difficulties of the morality of common sense in treating with hard cases without seriously discussing the philosophical efforts developed to bring consistency and systematisation to it, see A. Donagan, *Sidgwick and Whewellian Intuitionism: Some Enigmas*, «Canadian Journal of Philosophy», 7, 1977, pp. 447-465; repr. in B. Schultz (ed.), *Essays on Henry Sidgwick*, cit., pp. 123-142, and S. Cremaschi, *Nothing to Invite or Reward a Separate Examination. Sidgwick and Whewell*, in this issue.

indispensable to rational moral life. At any rate I thought I might provisionally postulate it, while setting out on the serious search for empirical evidence» (HSM, p. 467). Such empirical evidence should have come from the parapsychological investigations to which Sidgwick devoted much efforts throughout his life. In 1874 his hopes had probably already weakened enough to make the Kantian postulation unacceptable: in fact, in the first edition of *The Methods*, he already declares what we also find in all other editions, that is, that he could not

fall back on the Kantian resource of thinking myself under a moral necessity to regard all my duties as if they were commandments of God, although not entitled to hold speculatively that any such Supreme Being really exists "as Real". I am so far from feeling bound to believe for purposes of practice what I see no ground for holding as a speculative truth, that I cannot even conceive the state of mind which these words seem to describe, except as a momentary half-wilful irrationality, committed in a violent access of philosophic despair (ME 1, p. 471; cf. ME 7, p. 507).

The process of disillusion had been (almost) completed by 1887, when he writes: «I have been facing the fact that I am drifting steadily to the conclusion — I have by no means arrived at it, but I am certainly drifting towards it — that we have not, and are never likely to have, empirical evidence of the existence of the individual after death» (HSM, p. 466). Lacking any such evidence, Sidgwick seems to be totally bereft of reasons to accept Kant's postulation.

Sidgwick might have solved the dualism of practical reason only by questioning the rationality of egoism, which he clearly was not quite prepared to do. As he makes clear in a 1889 paper, such rationality is based on the reality and fundamentality of the distinction between any one individual and any other, so that «'I' am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals»[45]; it is in fact based on the very idea of the «separateness of persons» urged by Rawls against utilitarianism. For Sidgwick, this preference of private happiness to virtue, or general happiness, is just as much a dictate of reason as the proposition that my own good is no more important than the good of any other; for Kant, the demands of reason are in no way conditioned to the effective reconciliation of happiness and virtue, which must be postulated and hoped for, but cannot be the motive of action: morality teaches us how to become deserving of happi-

[45] H. Sidgwick, *Some Fundamental Ethical Controversies*, p. 44.

ness, without assuring that we will actually be happy. In the end, persons are for Sidgwick much more separate than they are for Kant. In fact, thanks to his identification of reason with the very capacity for universality, Kant can question the rationality of egoism; to be rational is in fact to acknowledge that maxims by which we are making exceptions for ourselves are not justifiable[46]. Persons are thus separate, but practical reason is a point of view with which all human beings can identify themselves, a common identity rooted in their nature as rational beings. For Sidgwick, instead, the rationality of overcoming egoism can only be seen by viewing things from the point of view of the universe; this, however, is actually no one's point of view: it is not a perspective rooted in our nature as the first person perspective that grounds rational egoism. This account seems to emphasize the separateness of persons more than the Kantian one, for here the viewpoint of universality seems not one that is shared by all, but one to be constructed by summing the perspectives of all; in this perspective, the concern with «the quality of my existence» cannot but trump any interest for universal benevolence.

In conclusion, there are some reasons that may suggest that a deeper understanding of Kant and a more direct confrontation with Kant's ethical treatises might have led Sidgwick to second thoughts on ethics, the rationality of egoism, and his final rejection of the deontological stance of the morality of common sense. It is a fact, however, that Sidgwick never embarked in such a confrontation: and my speculations on the possible reasons for this circumstance lead to single out both his distaste for Kant's metaphysics and his Millian utilitarian bias[47].

[46] Kant's dualism between instrumental and moral rationality is therefore very different from Sidgwick's dualism (and from other dualisms recalled by Sidgwick), as was shown by W. K. Frankena, *Sidgwick and the History of Ethical Dualism*, in B. Schultz (ed.), *Essays on Henry Sidgwick*, pp. 175-198.

[47] I wish to thank Sergio Cremaschi and Gianfranco Pellegrino for very helpful comments on an earlier draft.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 137-184

"Nothing to invite or to reward a separate examination": Sidgwick and Whewell

Sergio Cremaschi
Dipartimento di Studi Umanistici
Università del Piemonte orientale
sergio.cremaschi@lett.unipmn.it

Abstract

In this paper I discuss Sidgwick's reaction to Whewell's moral philosophy. I show how, to Sidgwick's eyes, Whewell's philosophy looked as an emblem of the set of beliefs, primarily religious, into which he had been socialised, and that his reaction was over-determined by both his own ambivalent feelings to his own Anglican upbringing and his subtle rhetorical strategy practised by presenting new shocking ideas hidden between an amount of platitudes and playing the neutral observer or the 'philosopher of morality' instead than acting the part of the preacher of a new morality. Then I discuss Sidgwick's assessment of Whewell's doctrine as an idle systematisation of received opinion and the reasons why in the *Methods* he feels entitled to dismiss historically given intuitionism as 'dogmatic intuitionism' without detailed criticism and discusses instead a so-called 'intuitional method' as one of the procedures allegedly used by common sense. Besides, I show how individual instances of detailed criticism to Whewell's doctrines are meant to be not 'real' criticism of a rival outlook but instead illustrations of the limits of 'common-sense morality'. My final claims are: first, Sidgwick ends with a short-circuit between a inner dialectic of his own argument and discussion of rival doctrines; second, the weight of Whewell's legacy in Sidgwick's ethics has been heavily underemphasized.

> *His elements of Morality could be nothing better than a classification and systematizing of the opinions which he found prevailing among those who had been educated according to the approved methods of his own country; or, let us rather say, an apparatus for converting those prevailing opinions, on matters of morality, into reasons for themselves…*
> *He leaves the subject so exactly as he found it…that it can scarcely be counted as anything more than one of the thousand waves on the dead sea of commonplace, affording nothing to invite or to reward a separate examination.*
>
> John Stuart Mill.

1. *Sidgwick and 'intuitionism': which and whose?*

"Probably nothing did more to discredit Whewell's system than Sidgwick's study of Intuitionism in his *Methods of Ethics*" (1). This is hardly surprising since, in a well-known passage, Sidgwick candidly mentions what he names "my early aversion to Intuitional Ethics derived from the study of Whewell" (2). In other words, it seems that the reasons for Sidgwick's strategy of dismantling Intuitionism and proving its irreparable limits was motivated by his antipathy to a book that had been a juvenile (compulsory) reading as well as to its author. In view of this circumstance, one may wonder whether Sidgwick's campaign has been so effective as to blur the memory of Whewell's ethics to the point that, until recently, the Sidgwick scholarship, while paying due attention to the topic of intuitionism in Sidgwick, did usually not go much beyond than repeating as a mantra the threefold distinction between perceptive, dogmatic, and philosophical intuitionism, and referring in all seriousness the information that dogmatic intuitionism was hopelessly unable to solve the dilemmas left by perceptive intuitionism and besides was a way of giving an appearance of intellectual respectability to moral prejudice. Some of the more recent literature tries to discuss the meaning and scope of 'intuitionism' in Sidgwick's ethics by careful textual reading and linguistic analysis of Sidgwick's own assertions, without even including in the bibliography the intuitionist authors whose views Sidgwick was criticizing or partially endorsing in the hope that real intuitionism is intuitionism as described by Sidgwick. The reasons? The usual ones, namely, Anglo-Saxon phobia vis-à-vis the history of philosophy, and world-wide spread powerlessness when facing the task of looking for books one cannot find in one's Department Library, besides the ruinous effect of Sidgwick's campaign.

The result is that everybody repeats, as if it were a source of objective historical information, what Sidgwick says in his preface to the seventh edition of the *Methods*, that is: he was disgusted by lack of clarity in definitions when compared with those by mathematicians (nothing less, with all that Aristotle has said about the lesser degree of certainty of the propositions with which practical philosophy has to start compared with the purely theoretical parts of philosophy) and he felt that this textbook he had to study as an undergraduate was a systematisation of all the unjustified moral teachings he had been imparted in his childhood. Nobody reflected about the circumstances that this was a senile restatement of events that occurred decades before; that these reactions referred to an item of compulsory reading in somebody's education; that this item was

signed by somebody who was one of the older dons of the same college as Sidgwick's, a generation with which Sidgwick had a conflictive relationship for many years; that this senile restatement echoes strangely Mill's opinion on Whewell's *Elements*, "nothing better than a classification and systematizing of the opinions which he found prevailing", that is, what the educated public had been in the meanwhile educated into thinking through the extraordinary influence won in the meanwhile by Mill as a "public moralist"; that Whewell had been a public figure in a context where he and Mill had been for a time the champions of the Old and the New, and that the New had won the war, so to say making no prisoner, and even in the Church of England the trend represented by Whewell had been wiped out and substituted by either a more 'progressive' trend – a kind of Anglicanised Unitarianism such as that proposed by Bishop Baden Powell (the father of George) and other liberal Anglican divines – or the more traditional trend of Evangelicalism, and last of all, that the strictly philosophical doctrines by Whewell were in Sidgwick's eyes not only intertwined with a wider overall view, religious and political, but were part of a set of beliefs (a moderately enlightened Anglicanism with a moderate liberal Whiggish political outlook) that were part of Sidgwick's own *Bildung*, that he never totally rejected and looked from outside but always cherished as a lost Ithaca to which he would have liked, were it possible, to come back some day.

This may be enough in order to account for ambivalences, turns, and tensions in Sidgwick's relationship to Intuitionism in general and Whewell in particular, but looking at these only, as Schulz tries to do (3), would only yield a 'genetic' history of ideas of one of the most familiar Continental kinds, and a not very enlightening one. What I suggest to do is instead taking this background into account and trying to detect which *things* Sidgwick was trying to do with *words*. That is, I suggest we should make the most of one remark by Schulz himself, namely that in his major works:

> Sidgwick appears to have applied the lessons that he had set out so many years before, for his friends in the Initial Society. That is, he became quite expert at masking the originality and subversiveness of his claims by the Mauricean tactic of presenting them as mere developments of received belief, cloaking his real insights with massive tomes of respectable opinion so that few could apprehend how destructive his criticism was trivialities… Perhaps, as with the *Methods*, Sidgwick always felt that the respectable views he criticized were enduring elements of his own

being, and that the criticism really was a form of self-scrutiny, an inner Socratic dialectic rather than "hostile criticism from the outside" (4).

I would like to add that Sidgwick staged a twofold strategy in order to deal with Whewell, a strategy indeed he tended to mount also in many other occasions: on the one hand he develops an inner Socratic dialectic with views that were still part of himself, albeit as a polarity of a Hegelian dialectic between beliefs we would like to have and beliefs we have to be rest content with, and at the same time he develops an external rhetorical strategy aimed at an audience made of a majority of educated and rather traditional Victorian readers and a minority of progressive Millian readers.

In my attempt, I start with conclusions reached by Donagan and Schneewind, the ones who first started reading the *Methods* as a text, not as an oracle. Donagan provided the proof of the rather obvious conclusion that Sidgwick had not really read Whewell's arguments on the main points on which he attacked intuitionism and that his refutation of intuitionist arguments is curiously enough a suggestion of the fact that common sense has no answer to a number of doubts concerning limitations in the scope of principles and conflicts among principles, not a detailed answer to arguments provided by Price, Reid, and most of all Whewell in order to settle the issues under discussion (5). Schneewind has taken a step further, namely he has read first Sidgwick not as Moore's reluctant stepfather, but instead in the light of the controversy between Mill and Whewell; in this way he has shown *why* to Sidgwick occurred the not-too-peregrine idea of reconciling utilitarianism and intuitionism and *where* he found the arguments in favour and against each of his own three methods (6). In Schulz's words,

an excellent way to approach the Methods is by reading it, as Schneewind has done, in the light of the great conflicts between Mill, the romanticized utilitarian, and Whewell, the intuitionist defender of orthodoxy whom Mill himself singled out as representing just about everything that utilitarianism should oppose (7).

In fact, in the former phase, the young Sidgwick found in Mill a spiritual guide in his own search for freedom of thought. It may be reminded that he corresponded with him at the time of his famous conscientious objection to subscription of the 39 articles of the Anglican faith required to Cambridge faculty members. Sidgwick mentioned later on also the cir-

cumstance that Mill's 'hedonism' sounded attractive to him as a kind of "relief from the apparently external and arbitrary pressure of moral rules" which he had been educated to obey (8). But in subsequent phases Sidgwick also discovered the attractiveness of a Goethian neo-pagan ideal and wavered more than once between the alternative enticements of mysticism, benevolence, and hedonism, or religion, utilitarianism, and romantic aestheticism. Also his way of reading more strictly philosophical doctrines was coloured by their associations with these more encompassing world-views. Also his way of reading Whewell's moral doctrine is overdetermined by his own personal experiences, that is, by the circumstances that Whewell, with whom he was directly acquainted, was to his eyes connected with the set of religious beliefs he had been imbibed with in his boyhood and to which he longed all his life long to come back, if only it were possible. Without such ambivalent personal experience, probably Whewellian intuitionism would have been discussed more at length and in a more detached manner, and the rather powerful dose of intuitionism Sidgwick finally thought it proper to take would have been openly acknowledged as Whewell's legacy. Finally, another factor played in favour of under-stressing Whewell's legacy, namely the wary rhetorical strategy-cum-tactic stages by Sidgwick. On the one hand Sidgwick as a public figure – the proponent of educational reforms, the women's rights etc. – had as partners both 'militant' Millians and respectable enlightened Anglicans; for both these groups the *Methods* were too obscure a work, and yet it was important not to arise polemics that could reach this wider audience; thus, not presenting himself too explicitly as an orthodox utilitarian was good for the latter part of his audience, not attacking too explicitly Whewell could have been good for the former, albeit at the time of the *Methods* Whewell's star was on the point of declining even in the Anglican firmament. Thus, a good tactic in order not to become either group's enemy could have been to pay lip service to Mill's attacks on Whewell, to present himself as being somewhere in between Utilitarians and Intuitionists, to keep silent on Whewell the rest of the time, and especially keeping up being rather tedious and obscure in the highest degree all the time.

2. *Whewell's philosophy of morality*

Let me come back briefly to Whewell's own ethics. It popped up, at last as a sketch, in his Preface to the 1835 edition of MackIntosh's *Dissertation*, and by 1845it was developed into a bulky work, the *Elements of Mo-*

rality (9). The work was written in order to provide an alternative to Paley's *Principles* that were still basic reading for undergraduates at Cambridge and whose negative influence had been denounced in Sedgwick's *Discourse* of 1832 (10). In order to provide an alternative to Paley, Whewell wanted to offer an anti-empiricist moral philosophy, well-tuned with his own anti-empiricist epistemology, rescuing ideas that had emerged in Cambridge Anglicanism at the end of the seventeenth-century and the beginning of the eighteenth but that had been totally wiped out by the Gay-Brown line of voluntarist consequentialism that was later systematized in Paley's *Principles* of 1785 (11). By doing so, the moderately liberal Anglican Whewell pillaged also the work of the Dissenter Richard Price, for rather obvious reasons, without stressing too much the circumstance.

It is fair to say, yet, that Whewell added a lot of his own, primarily a para-Kantian moral epistemology, which made room for an a priori element in moral discourse while making it compatible with varying historical institutions by a sort of 'circular' development that provides the blueprint for human knowledge, both in the natural science and in ethics: from facts to principles and from principles to facts and another quasi-Kantian idea, the idea of a 'fact' of moral judgement that needs clarification but does not require any justification. Whewell's epistemology turns around the idea that empiricism heads to vicious circles, and this idea was more or less at the centre of his first controversy with Mill, concerning induction (12). In ethics too Whewell contends that empiricism, like in Paley and Bentham's case, heads to a vicious circle, or a hopeless tangle made of virtue and happiness. Against empiricism, he defends an idea of ethics as being indeed a science – what the empiricists agreed on – but a science of a peculiar nature, aiming at some objective truth that is a *specifically moral* truth – a point on which he parted company with empiricism. Yet one idea he has clear in mind – it is worth stressing it when facing Sidgwick's criticism to intuitionism – is that we do not need to assume that we already possess it in full, but it may be a kind of truth we acquire step by step, not unlike what happens in the natural sciences (13), whose development follows a spiral-shaped pattern travelling between two opposite poles, namely clarification of the Idea and discovery of Facts. In both physics and morality,

> all truths include an Idea and a Fact. The Idea is derived from the mind within, the Fact from the world without (14).

Not morality, but a "philosophy of morality" is the philosopher's subject, since the former already exists, and may be recognized even when the eventual reasons for its justification are still a matter of controversy, not unlike the theorems of geometry which are agreed upon by mathematicians who disagree in their philosophies of mathematics. This philosophy of morality combines Ideas and Facts trying to build a deductive system, which can absorb results of previous systems but be more consistent, eliminate inner contradictions and inability to account for *moral* facts. The latter are particular evaluative or prescriptive judgments that present themselves as undeniable to everybody's conscience. *Brute* facts are the laws enforced in one society, viewed at within the framework of the process that made them such as they are and accordingly, "though we have, in different places, different Laws, we have everywhere the same Morality" (15).

Existence of moral facts is proved by the existence of public opinion or by "the great fact of the universal and perpetual judgment of mankind on actions as just or unjust"(16), from which a lesson may be drawn, namely the fact that

> man cannot help judging of actions, as being right or wrong; and that men universally reckon this as the supreme difference of actions… this characteristic of human nature marks man as a moral being; as a being endowed with a faculty or faculties by which he does thus judge (17).

And this fact is indeed "the beginning of all morality" (18). Whewell does not claim that "this *Faculty* or those Faculties by which man thus judges of right and wrong should be anything peculiar and ultimate, but only that the *distinction* should be a peculiar and ultimate one" (19). It is in so far as human beings form such judgements, not in so far as they feel pleasure and pain that they are moral creatures. These, unlike the facts of natural science, are prescriptive facts, consisting in the whole of the norms imposed by the laws and the public opinion of one society to its members; this is the prescriptive form of what a society assumes to be moral facts.

The moralist's task is working out a set of "Ideas" that will account for these facts as a whole, while occasionally correcting their account on individual points. In other words: morality qua phenomenon is a fact; ethics as an intellectual discipline consists in a twofold task: first, providing a rational reconstruction of morality qua phenomenon, second, working out a philosophy of morality, that is a clarification of the ways moral-

ity works and of the grounds for its justification. A preliminary task for the philosopher is working out a consistent account of the contents of morality so that philosophical reflection may become possible about a well-defined subject matter. In the Preface to the first Edition of the *Elements* Whewell declares that

> Morality and the philosophy of Morality differ in the same manner and in the same degree as Geometry, and the Philosophy of Geometry… Men would never have discussed whether and why Geometrical Truth was possible, if they had not had before them and undeniable collection of such truth. Or, if without having any certainty or knowledge of Geometrical propositions, Men had speculated and disputed, as to whether they could have such knowledge and such certainty; we cannot suppose that they could have arrived at any distinct or stable result of such speculations (20).

The current distinction between metaethics and normative ethics is believed – fairly enough – to date back to George Edward Moore's formulation of a distinction between 'ethics' and 'casuistry'. Yet, it is clear enough that an idea of ethics as purely theoretical discipline, distinguished from normative ethics is already present in Sidgwick's often quoted anti-Aristotelian dictum "not practice but knowledge" (21). What is less known to Sidgwick's readers, but was indeed quite clear to Sidgwick – is that Whewell had introduced a distinction between Morality and the philosophy of Morality on whose basis the latter became a purely theoretical science, and besides that the construction of a consistent system of morality was a prolegomenon to any fruitful discussion of theoretical issues concerning the nature and justification of ethics (22).

3. *Whewell's system of morality*

The leading idea in our search for true moral propositions is that man acts qua man only when he acts under the guidance of reason, and the latter addresses us towards norms; the latter thus are required for the action of man as a man; indeed we cannot conceive of man without conceiving him as subject to norms and belonging to some norm-based order (23). The proof lies in the fact that the authority of reason over our desires is self-evident, for man is seldom impelled merely by the most elementary springs of action, bodily desires and affection" (24) but most of the time,

they "are unfolded by thought, so as to involve abstract conceptions and the notion of a Rule" (25), and in case of conflict between desire and reason, we are aware that that our own act is the one we carry out in accordance with reason, and the reason for this is that "the Reason alone is capable of that reflex act by which we become conscious of ourselves" (26).

Our quest for a set of moral truths leads us to five basic ideas, implicitly underlying all the moral facts we discover by observation of law and custom in different societies, and besides to a basic moral norm that turns out to be a fundamental principle or axiom of morality as such. The five ideas of benevolence, justice, truth, order, purity "are dispositions conformable to the Supreme Law of Human Action: they are Virtues" (27), and they provide specific contents to the "Supreme Norm of morality". The latter may be described by its end, that is the True End of human action or the *Summum Bonum*, and may be framed in several alternative ways, such as "we *ought* to do what is right; we *ought not* to do what is wrong. To do what is right is our duty; to do what is wrong is a transgression, an offence, a violation of our duty" (28).

A need for a Supreme Norm arises out of the need to answer questions about the justification of particular norms. The succession of means and ends with a corresponding succession of subordinate and superior norms has to stop somewhere. Thus, concerning the Supreme Rule, the question "why?" admits of no further answer. "Why must I do what is right? Because it *is* right. Why should I do what I ought? Because I ought. The Supreme Rule supplies a reason for that which it commands, by being the Supreme Rule" (29).

Whewell's claim was that morality arises from the *Intellect*, not from *Sense* (30). Only in the Preface to the second edition a concession is made to common sense, the notion cherished by his Cambridge idealist colleagues. He writes:

> Morality has its root in the Common Nature of man; and no Scheme of Morality can be true, except a scheme which agrees with the Common Sense of mankind, so far as the Common Sense is consistent with itself: including in the term Common Sense, both men's convictions as to what is right, and their sentiments as to what is morally good (31).

Whatever Whewell's intentions in making such a concession, the fact is that rules of morality are derived from the Supreme rule and the binding character of the latter lies in its character of an axiom. That is, com-

mon sense cannot but confirm rational morality, but the latter does not need the former in order to be justified.

A serious traditional problem for which Whewell undertakes to provide an answer is the possibility of a conflict of duties. Whewell suggests that such possibility has been too much emphasized by casuists in order to find excuses for the omission of duty itself. A real conflict between duties arises only in case of "extreme necessity", while in the majority of cases of necessity there is an *excuse* for transgressing the moral law, but not a *real conflict* of duties (in these cases one could avoid to transgress the moral law and sacrifice one's life as a heroic act, which would be supererogatory). There is genuine conflict between duties only

in the case in which Moral Rules are transgressed, not for the sake of our own preservation, but in order to preserve *some other person* from great impending evil; we may have a Case of Necessity, which is also a *Conflict of Duties*: for to preserve another person from great evil, is a part of the general Duty of Benevolence; and when the person is connected with us by special relations, to do this, is involved in the Duties of the Specific Affections (32).

Only in such cases "we have two Duties, placed in opposition to each other; on one side, the Duty of rescuing, from a terrible and impending evil, a husband, a friend, a daughter, a neighbour; on the other hand, the Duty of not telling a falsehood, or committing homicide" (33). For such cases "the Moralist must abstain from laying down definite Rules of decision" (34), firstly because in such cases a previous decision is difficult and accordingly general rules are of little use. Besides, to state

General Rules for deciding Conflicts between opposing Duties, would have an immoral tendency. For such a procedure would necessarily seem to make light of the Duties which were thus, in a general manner, postponed to other Duties; and would tend to remove the compunction, which any Moral Rule violated, ought to occasion to the Actor (35).

It is unavoidable that law be violated, but it is a good thing that compunction is left; the moralist's task cannot be teaching the lawfulness of violating the law. People in cases of necessity will have no time to consult the rules laid by the moralist, but "they will be determined in their conduct on such emergencies, by their previous moral culture and moral progress (36). Such cases are indeed real occurrences, and virtues dis-

played in such cases are on the same occasion called heroic virtues, since tragic choices depend on a too strong adherence to one moral principle. Yet they may be admired to a point but not be recommended for imitation, since to aim "at Heroic Virtues only, would be an extremely bad culture of ourselves. It would lead to an entire rejection of Duties" (37).

Whewell's main point yet is that moralists have overemphasized the possibility of conflict of duties. Most of such conflict is apparent one, since they simply arise from the existence of a plurality of principles, not by cases where danger of death is impending on some person to which we have duties of affection. Mere coexistence of conflicting rules creates indeed problems, but such as may on principle be settled by rational argument and problems concerning not such a disturbing question as "How may Duty be evaded?" but a more plain question such as "What ought I to do?" (38). The most typical of such questions, addressed in ch. 15, is veracity, or keeping promises and telling the truth, a matter discussed by moralists for centuries and about which a few quite questionable conclusions have been circulated as if they were respectable opinions. Whewell's general line of argument is that in most cases there is no need to ask whether we may be dispensed from doing what is our duty, since there are doubtful cases where it may be proved that it is or it is not our duty to keep a promise or to tell the truth. The general premiss is that words are not to be understood literally but according to the "mutual understanding" which the use of language implies (39). From this general principle in several cases the proof may be given that one has no duty to keep a promise because a mutual understanding concerning the truth of a number of conditions is implied in every act of promising; this is why I have no duty of fulfilling a promise in case that "the Common Understanding of what the Promiser is to do for the Promisee, includes some false suppositions which are afterwards discovered to be false" (40). Whewell's settlement of the allegedly doubtful case is that "the false supposition releases the Promiser, so far as it was included in the Common Understanding" (41. On the basis of such general principle Whewell gives an answer for a number of traditional debated issues, and on three specific cases argues an answer more rigorist than Paley's. These are the case of the promise extorted by fear, where he argues that the promise, if morally made, should be kept, even in cases where the law allows for duress as an extenuating circumstance. It is worth noting that Whewell argues that, even taking consequences into account, these are so uncertain that they can hardly play in favour of one alternative; for ex., will not paying a ransom dis-

courage hijackers from further kidnapping, or will it prompt them to "add murder to robbery?".

Even on the balance of probable advantage, it would seem that such a promise is to be kept.

But on our principles, we should not look to these results as to our own moral culture. By keeping this promise, we cherish and exemplify our regard for truth. What moral quality do we cultivate by breaking it? If it be replied, that we thus cultivate a regard for consequences; we reply, that consequences, when both their existence, and their moral character are so doubtful, are not the main objects for our regard (42).

Another case is the one of the author of an anonymous work who, according to Paley, may deny his authorship while, according to Whewell, may try to guard his secret by avoiding to answer by various devices, but cannot tell overtly a lie, for all he may suffer is "some vexation or inconvenience", while by succeeding in keeping his secret at the expense of truth "he receives a moral stain" (43). Another case is that of lies told by advocates in favour of their clients, admitted of by Paley and ruled out by Whewell (44). One more is the promise made to a woman by a married man to marry her in case his wife would die. Paley's answer was that it is wrong to claim that the promise was void "for, however criminal the affection might be, which induced the promise, the performance, when it was demanded, was lawful; which is the only lawfulness required" (45). Whewell's more complex answer is that, even if the promise is immoral, and by implication void, the duty to marry the woman does not depend on the immoral promise alone and the promiser may marry her since the promise "does not necessarily vitiate all the succeeding dispositions to the woman to whom the promise was made" (46).

The allegedly dubious cases thus settled differ from one case, where the same dilemma presents itself for truth as for any other duty; this is the case of extreme necessity, where what is at stake is not some inconvenience but life itself, or, even worse, not the agent's but that of a third person's life. Here, as in all similar cases, a breach of duty is excusable in the former situation, and is even required in the second, in so far as, by carrying out a lesser duty, we would violate a heavier one.

Besides truth, also justice – discussed in ch. 21 – may be a ground for (real or alleged) conflict of duties. Rights are a condition for man's action; they are defined by the State; but there is widespread a fundamental conviction, that rights are arbitrary. In other words, there is Natural Law,

depending upon the nature of man. Such law is not found somewhere else than in existing systems of law, and yet it is not coincident with any of them. The solution to the apparent dilemma arising is that

Right cannot be founded on Injustice: such is the negative maxim, which serves to define the Idea of Justice. *Justice assigns Rights according to existing Conditions*: such is the positive maxim, which makes Justice applicable to facts (47).

That is, there is an ideal and an arbitrary element in any legal framework of rights. It is positive law that assigns specific rights, such as those of property, and in doing so it depends on facts, that is on "circumstances, which are not governed by our Ideas" (48), but existing arrangements should be constantly improved in order to bring them more and more close to requirements of justice. How much and when is matter of external circumstances, and cannot be dictated by the Ideal element, but the idea of a Natural Law does not consecrate existing arrangements, on the contrary provides a standard for amending them.

Sidgwick's criticism to intuitionism in the *Methods* will concentrate precisely on these two points, truth and justice, assuming that they are the paramount cases where the inability to settle dubious cases by the 'method' of intuitionism is particularly apparent.

4. *The Mill-Whewell controversy*

The controversy on ethics between Mill and Whewell took place between 1852 and 1854, following another on philosophy of science, more precisely on induction. The difference between the two controversies is that the former was more academic in tone and in its course Mill paid due respect to Whewell's superior merits in the field of the history of science, the latter had all the aspects of a public controversy, one where what is at stake is control over the public opinion and the ultimate issue is, rather than a theoretical one, who is going to be the ruling group in a given society at a certain historical phase (49).

A start was provided by Whewell's criticism to Bentham in his *Lectures on the History of Moral Philosophy in England* of 1852 (50). Mill thought it proper to attack openly Whewell after he had published an explicit criticism of Benthamite ethics, probably in order to be in a stronger position than if he had criticized the *Elements*, since he was in position to

complain of the fact that utilitarian ethics had been unfairly misrepresented. Mill was not new to such exploits. A good example is his previous attack on Sedgwick's allegedly "intemperate assault on analytic psychology and utilitarian ethics, *in the form of* an attack on Locke and Paley" (51). Sedgwick had criticized Paley without even mentioning Bentham, and was made the target of Mill's vehement counter-attack starting with the curious proviso that he Mill would not spend a word in defence of Paley, since he was a priest and hence a preacher of reactionary ideas. Mill's odd argument is that, since Sedgwick, while criticising the reactionary and superstitious Paley had *implicitly* attacked Benthamism for what the latter shared with Paleyism, and therefore he was twice guilty, for having attacked (implicitly) utilitarianism and for having ignored it (explicitly). The reason for Mill's choice in this case was the – very good one indeed – that Sedgwick's *Discourse* had enjoyed an enormous circulation and could accordingly grant comparable popularity also to its critic. Also in the case of Whewell's *Lectures* the reason for the attack was the author's prestige, besides the fact of having offered an occasion for *complaining* of something, misrepresentation, unfair criticism, bad faith in attacking a doctrine just because it subverted established prejudice etc., instead of expressing sentiments of gratitude for the fact that an established intellectual authority had dedicated no less than 63 pages to a discussion of the (until then neglected or at best execrated) Benthamite ethics (52).

The XXI century reader might ask why rationalism should find itself siding with religion, tradition, and political conservatism, while empiricism in turn was taking sides with atheism, progress, and political liberalism. Mill's reasons were the following:

> the notion that truths external to the mind may be known by *intuition or consciousness*, independently of *observation and experience* is, I am persuaded, in these times, the great intellectual support of false doctrines and bad institutions. By the aid of this theory, every inveterate belief and every intense feeling, of which the *origin* is not remembered, is enabled to dispense with the obligation of justifying itself by reason, and is erected into its self-sufficient voucher and justification (53).

But such an account sounds slightly odd. After all, on one hand, Edmund Burke, the most able advocate of traditionalism, had based his own argument precisely on anti-rationalism, Hume, an empiricist if any, defended a kind of mild Toryism; on the other, William Godwin had been a

rationalist radical, Richard Price, the intuitionist moral philosopher, had been a supporter of the cause of American independence and had been attacked by Bentham from a more moderate stance.

At the time Mill wrote his own attack on Whewell *intuitionism* was comparatively a novelty, and the very word intuitionism as the name for an ethical doctrine arouse out of Mill's own classification of ethical thinking into the empirical school and the "doctrine of intuitive principles of morality" (54). There had been indeed a rationalist tradition in British ethics from the end of the seventeenth century, but in its first phase it was more Platonic than intuitionist in Whewell's sense. In fact their main claim was a kind of moral realism, that is, a thesis in moral ontology, not a thesis in moral epistemology. Towards the end of the eighteenth century the only advocate of some kind of 'intuitionist' ethics (as far as he put forth a claim in moral epistemology, namely that there are a number of prescriptions that cannot be denied at the price of logical contradiction) had been Richard Price. The Scottish school, Thomas Reid and Dugald Stewart, defended against Humean empiricism the existence of moral principles belonging to the common sense, not to our rational faculties, which is in turn a peculiar claim, different from both moral Platonism and moral intuitionism as I have defined it. As a consequence, one may wonder who were the enemies Mill wanted to fight in his youth, since there was hardly any rationalist or apriorist school around at that time defending both erroneous doctrines and bad institutions. The Scottish followers of Dugald Stewart were outsiders to the establishment and committed liberal reformers, less radical than the Benthamites, but still clearly fellow-travellers, not enemies. Cambridge had been Paley's own preserve, and Mill manifested despise for Paley and his followers – their empiricism notwithstanding – because of their conservative position cloaked under progressive language. and thus one may wonder why rationalism, should be blamed for all the evils existing in the world.

Whewell in his *Lectures* argued that Benthamite moral theory was defective on two main points, namely the impossibility to calculate all consequences of actions and the circumstance that happiness includes moral elements, and thus we cannot properly derive morality from happiness. Let me illustrate this criticism in more detail. Whewell wants to rule out the claim that morality be a means to some end, which in turn is not moral in its nature (55). He concedes for the sake of the argument the truth of the assertion that "acts are virtuous in proportion as they calculably produce happiness" (56) if we take all acts as a whole into account

and calculate all consequences, but he argues that, even on this premiss, it turns out to be impossible to make this assertion the very basis of morality. The first, already mentioned, reason for this impossibility is our inability to calculate all consequences of an act, or to solve so difficult a problem as that of establishing, among two lines of action, which one will yield the maximum amount of happiness (57); there is yet a more simple way of deducting such rules, that is, as Whewell explains in the Preface to the Second Edition of the *Elements*, considering that human beings living among other human beings need such rules and, "by the mere contemplation of our human faculties and springs of action, we can discern certain relations which must exist among them, by the necessity of man's moral being" (58). The second reason for this impossibility is that happiness includes moral elements, and thus we cannot just derive morality from happiness without falling into a vicious circle. For example, we may ask, "Why should a man be truthful and just? Because acts of veracity and justice, even if they do not produce immediate gratification to him and his friends in other ways… at least produce pleasure in this way; that they procure him his own approval and that of all good men" (59). This may be all right, but a Benthamite would add that he "thinks it virtuous, because it gives him pleasure: and it gives him pleasure because he thinks it virtuous. This is a vicious circle" (60).

In 1852 in the *Westminster Review* Mill attacked Whewell's *Lectures on the history of Moral Philosophy* and his *Elements* in a long essay. He explained that he had not discussed in public the *Elements* until the *Lectures* too were published, because the former work was of limited interest as such – being a "mere a catalogue of received opinions, containing nothing to correct any of them, and little which can work with any potency even to confirm them" (61), and that he felt that a rejoinder was required after Whewell's attack on Benthamite doctrines in the *Lectures*, and finally that he felt that a consideration of at least some parts of the *Elements* was needed in order to expose the roots of Whewell's mistake (62). Mill argues first that Whe well in epistemology and ethics adopts arguments that justify use of a priori theses not derived from experience and in this way he finds a theoretical argument for justifying the transformation of the precepts of traditional morality in a system of allegedly self-evident truths (63). Then he argues that Whewell's stronghold, that is his idea of a fundamental norm, that we must do what is right, is a tautology and thus does not contribute anything positive, unless we admit that doing what is right is equivalent to not violating rights, and in this case his sys-

tem of morality is made dependent on positive law, so that his rule of right is

> to infringe no rights conferred by the law, and to cherish no disposition which could make us desire such infringements! According to him, the early Christians, the religious reformers, the founders of all free governments… and all enemies of the rights of slaveowners, must be classified among the wicked (64).

Thirdly, he claims that to make morality depend on other elements, themselves moral, as Whewell wants to do, would end up with a vicious circle, but actually Whewell cannot keep up his own standard and in the end he admits that morality serves other ends, themselves not moral in their nature, that is preventing "a disturbed and painful state of society" (65), but – Mill comments – this is utility or, in a word, when "real reasons are wanted, the repudiated happiness-principle is always the resource" (66).

Whewell responded to Mill's criticism in ch. 2 of the bulky Addendum he wrote for the third edition of the *Elements*. The points he made are: first, that his reasoning was not circular because *right* means just "what must be done", and there is no further reason, that is, no "why" is introduced for doing what must be done (67); second, that he was not a utilitarian in practice, since he did not *derive* fundamental rights from human happiness, even while agreeing that they serve also this purpose (68); third, that he had not based morality on law, but just used law as an "indication of its place and form" (69).

To sum up, in Snyder's words, Whewell claims there is

> a progressive intuition of necessary truth in morality as well as in science. Hence it does not follow that because the moral truths are axiomatic and self-evident we currently know them… Nevertheless, Whewell does claim that we can look to the dictates of positive law of the most morally advanced societies as a starting point in our explication of the moral ideas. But he is not therefore suggesting that these laws are the standard of morality… Mill is therefore wrong to interpret Whewell's moral philosophy as a justification of the status quo or as constituting a "vicious circle." Rather, Whewell's view shares some features of Rawls's later use of the notion of "reflective equilibrium" (69).

And, if we try to assess the *reason d'être* and the outcome of this controversy, these turn out to have been rather bizarre ones. Looking at the philosophical pulp, not at the political rind, they were both trying to do something that was, albeit not the same thing, at least something that was much closer than Mill realised. I may conclude, again in Snyder's words, that

Their conceptions of morality were quite similar in some important respects. Both men eschewed the utilitarianism of Bentham, which asserted that pleasure was the sole determinant of virtuous action. Instead, both erected moral philosophies that stressed the importance of creating morally excellent characters that would find happiness in acting virtuously. Both believed that a proper education – one aimed at "cultivating minds" – would help in creating this kind of moral character. Moreover, both had hopes that a widening of the scope of this type of education could lead to an improved society (70).

5. *Sidgwick's* Holzwege: *from morality towards religion, heading nowhere*

The impact of this controversy on Sidgwick could be hardly overemphasized. The two decades after it were his formative years, and he struggled hard in order to find his own intellectual path. What should be kept in mind is that Mill was the winner on the 'external' ground: in the following three decades that "marked the peak of Mill's reputation and influence as a public figure, and he quite deliberately set about exploiting his acknowledged intellectual authority to promote certain social and political views as they related to the leading public issues of the day" (71), and while Whewell was being rather quickly forgotten by the academic and even the religious establishment. On the other hand, I would dare to suggest that Whewell was in a sense a winner of the controversy on the 'inner' ground, in so far as several of the changes and qualifications to Benthamite utilitarianism introduced by Mill in Utilitarianism of 1861 were precisely on points raised in the controversy with Whewell, and – let me add – while paying lip service to Bentham and manifesting execration for the "intuitional" moralists, incorporating much of Whewell's criticism into his own revised version of Utilitarianism (72). But what happens on the inner ground is of interest only to academic scribblers, while what happens on the open battlefield determines who is going to be the boss, which books will be reprinted, which books will be adopted in universi-

ties, which names will be mentioned reverentially by semi-educated elites, and in fact the fourth edition of the *Elements* of 1864 will be the last one for one and half a century while Mill's *Utilitarianism* will be reprinted and translated into many languages an incredible number of times, and authors of textbooks in many countries have been repeating just what Mill said about intuitionism.

In his different phases, Sidgwick kept on being, as a whole, a follower of Mill, at least on things that really matter, that is, everything but philosophy. He wanted to find his own tortuous path to truth, at times defining himself a utilitarian and at times leaning towards Kant and Butler and Reid, or alternatively towards Goethe and perhaps the Greek philosophers. But in the phases in which he looked for intuition, as against empiricism, he was careful in styling himself as a critical follower of the progressive party, leaving as little room as possible to suspicions of sympathies for the establishment, old Cambridge, and the Church of England. This is why he chooses his allies, in this phase, in Germany or in the British eighteenth century. Besides, he depicted himself on purpose as an impartial inquirer into truth in moral matters, a scientist, as contrasted with a preacher. He even added, while actually recalling Aristotle's project of transforming common-sense morality into a consistent system of opinions, a kind of Spinozean flourish in declaring that also the study of morality may be undertaken not in order to become better men – as Aristotle believed – but just for love of truth, like the theoretical sciences, a view that would be incompatible with Aristotle's view of practical philosophy as different in goal and standards from theoretical philosophy. Sidgwick also referred to his own encounter with the *Nichomachean Ethics* as some kind of revelation of the right kind of job to be carried out by moralists, but nonetheless, one page before, he mentioned "mathematicians" as embodying the standard of precision and clarity on which the "Intuitional moralists" should be judged (73). This self-image – as argued by Schultz – has much to do with his own rhetorical strategy, which may be summarized as follows: present a few subversive ideas on top of a ballast of shared opinions, mix heresy with Philistine common-sense, call all this 'science' or 'philosophy' and vindicate freedom of speech in the name of the impartial and objective approach you are entitled to adopting in so far as you are a member of the elite and a professional philosopher.

On the contrary, ethics was for Sidgwick a subject with deep existential implications, verging even more than on practical morality on the issue of the meaning of life, the existence of design and purpose in the world, and the problem of evil or theodicy. Sidgwick's real problem was

whether there is a way of reconciling rationality with the set of beliefs into which he had been educated and to which he had preserved a deep attachment up to the time of his studies at Cambridge. Such set of beliefs de facto meant Anglican theology of a non-traditionalist as well as non-Evangelical kind and the rationalist ethics taught by Whewell. Had he been a Cambridge undergraduate three decades before there would have been Paley's consequentialist voluntarism instead of Whewell's rationalist intuitionism. As a matter of fact, since the constellation of elements he had to face was this one, Sidgwick's idea of a philosophical defence of traditional morality amounts to Whewell's rationalism, and he seems not to be aware of the fact that the very same set of precepts had been taught for centuries cloaked under a Thomist, an Aristotelian, and more recently in England a consequentialist voluntarist philosophical jargon (or rather, he seems to refuse to draw consequences from something he knew too well). He wrote, at the time of the sixth edition, that as a teenager he felt uneasy under

> the apparently external and arbitrary pressure of moral rules which I had been educated to obey, and which presented themselves to me as to some extent doubtful and confused; and sometimes, even when clear, as merely dogmatic, unreasoned, incoherent (74).

He added, that his feelings of uneasiness were but

> intensified by the study of Whewell's *Elements of Morality* which was prescribed for the study of undergraduates in Trinity. It was from that book that I derived the impression – which long remained uneffaced – that Intuitional moralists were hopelessly loose (as compared to mathematicians) in their definitions and axioms (75).

Did he remember – while writing so – what Aristotle had said about different degrees of precision admitted of by theoretical and practical philosophy? Apart from this, the reported version of the story is something Sidgwick wrote thirty years after the first edition of the *Methods*. A circumstance worth stressing is that Whewell's book was the textbook he had to study as an undergraduate, that his feelings to it may have been over-determined by the way he felt with regard to his own previous moral education. It is far from clear that Sidgwick ever read seriously the work at a later stage when he discussed the "Intuitional moralists" in the *Methods*, and the impression may be not unjustified that, for various rea-

sons, he did not. One of these reasons may have been that he believed that it was necessary to distinguish between intuitional "ethical writers… who have confined themselves mainly to the definition and arrangement of the Morality of Common Sense, from those who have aimed at a more philosophical treatment of the content of moral intuition" (76) and that "the more philosophical school is the earlier" (77), that is Clarke and for some aspects also Butler. Another reason may have been that he was interested in intuitionism more as a possible 'method' he partly shared and this kind of intuitionism was a way of dealing with, and improving, common sense, and accordingly he was more interested in what the Scottish common sense philosophy had to say than in what Price and Whewell, the real intuitionists according to my definition of the term, had to say, and in facts he seems to ignore totally the former and to repeat on the latter the judgment passed on him by Mill, that of being the author of a "classification and systematizing" of moral prejudice, without apparently having taken the most theoretical part of the *Elements* into serious consideration. A third reason is that he did believe there were no serious discussions of ethical issues by 'really-existing intuitionism' and accordingly did not examine such discussions in detail preferring to concentrate on his own home-made intuitional method or on the conclusions allegedly reached by "Common Sense Morality", which he sought elsewhere, in writings by jurists or in prevailing opinions as he was able to reconstruct them through amateurish sociological observation. That is, as Donagan aptly remarked, there is a qui-pro-quo in Sidgwick's confrontation with Whewell and the intuitionists in general, arising from his assumption that, in order to be able to vindicate a self-evident character of moral first principles one should assume that morality be already evident in all its implications to common sense (78). Such request is too demanding and fails to meet Whewell's explicit argument that "in *moral* no less than in *physical* speculation", we face "a gradual and successive clearing and unfolding of those ideas which, on each subject, our knowledge must include, and in terms of which those speculative truths at which we arrive must be expressed" (79).

And yet, even if one could hardly believe that Sidgwick could lapse into such a blunder, an explanation of the reasons why he actually did could be found in his own overall strategy.

Before discussing this strategy, let me add something on the horizon of existential questions within which his inquiry into the so-called 'methods' of ethics took place. Sidgwick oscillated at different times of his life between Millian empiricism and some kind of mysticism, and between

'Christian' ascetism and a pagan or romantic experientialist approach to life. He wrote that among "the deeper problems" in which he was interested at the time, the main one was that of reconciling his "religious instinct" with his "growing conviction that both individual and social morality ought to be placed on an inductive basis (80).

The following year he wrote:

I am revolving a Theory of Ethics… I think I see reconciliation between the moral sense and utilitarian theories (81).

And shortly after he added:

My instinct for it [mysticism] is yet so strong that I am gradually developing my intuitive theories… You know I want intuitions for Morality; at least one (of Love) is required to supplement the utilitarian morality, and I do not see why, if we are to have one, we may not have others. I have worked away vigorously at the selfish morality, but I cannot persuade myself, except by trusting intuition, that Christian self-sacrifice is really a happier life than classical insouciance… That is, the question seems to me an open one. The effort to attain the Christian ideal may be a life-long painful struggle; and therefore, though I may believe this idea when realised productive of greater happiness, yet individually (if it is not a question of life or death) my laxness would induce me to prefer a lower, more attainable Goethean ideal. Intuitions turn the scale. I shall probably fall away from Mill and Co., for a phase… Another way out of it is finding the foundation of Christianity inexplicable by ordinary laws, and therefore, as the *vulgus* [do], worshipping the mystery, and obeying (child-like) the moral and religious intuitions of Christ, and, to a certain extent, of the Apostles (82).

If we look at Sidgwick's swings between different 'methods' through these letters, the different 'methods' of ethics start looking less as purely logical possibilities open to the human mind, and more as real-world alternatives. The choice among such alternatives had little to do with disinterested speculation, if the slogan "Knowledge not Praxis" is understood according to the prevailing mood of mainstream analytic philosophy, made of technical refinements, discussions of purely academic issues, and avoiding the Big Questions, or instead it was precisely 'disinterested' speculation of the best kind if one understands by the word open-

mindedness in a quest for the meaning of life. Basically, Sidgwick, the son an Anglican Rector (not unlike Nietzsche, the son of a Lutheran Pastor, and Durkheim, the son of a Rabbi) mourned until the end of his life over the death of God, longed for his resurrection, and found it over and over again impossible; in the while, he had found a substitute for his lost Ithaca in a progressive and humanitarian movement, Millian utilitarianism, not unlike Durkheim in France became an adept of republicanism. Both Sidgwick and Durkheim illustrated *ad abundantiam* the shortcomings of their respective secular churches' Creed, but also argued that people should be made to believe in such assumption as a token for non-existing more grounded ideals. Nietzsche took a different turn when he denounced humanitarian secular churches as the last harbingers of superstition, and looked bravely for the coming of some kind of ultra-man, one that could do without humanitarianism and pseudo-churches. Coming back to Sidgwick, it is as well to quote Keynes's famous dictum according to which he "belonged to the tribe of sages and pastors" (83) and elaborate on Keynes's suggestion, speculating that perhaps he *wished* he still could be a Christian and, precisely because he knew too well this was impossible, he *wanted* to be a Millian. He never betrayed – his mixed feelings to Mill himself notwithstanding – his loyalty to the Millian camp, not so much on theoretical as on real-world issues, precisely because, after the loss of the Christian faith – a Millian 'Religion of Mankind' was everything he had to preach.

In his unfulfilled wish to be a Christian, the great question Sidgwick kept on asking himself was one not infrequently asked in the nineteenth century, first by Kant and then in England by Coleridgean Idealists, namely, after we have proved that some kind of moral order in the human world has its own justification (the typical Enlightener's claim), is there a way to travel from the assumption of a moral order in the human world to a different claim, that of a cosmic moral order, implying the existence of a God as a judge? To this question, another – also a legacy of the Enlightenment – was added, namely, why is there underserved evil in the world? These questions were the ones debated by the Cambridge idealist sympathisers of Coleridge, first Frederick Denison Maurice and then George Grote and a number of less known figures with whom Sidgwick had been in close contact for decades (84). These were the questions that really mattered for Sidgwick. In a sense *The Methods* is a more an essay in theodicy than a treatise of ethics. In 1888 Sidgwick declared that "somehow or other, morality will get on" (85) and that maintaining morality

not *somehow* but establishing it "logically as a reasoned system" was an impossible task if we were to admit that

> we are limited to merely mundane sanctions, owing to the inevitable divergence, in this imperfect world, between the individual's Duty and his Happiness (86).

The dismal conclusions of the first edition of the *Methods* – the two very last words were "unavoidable failure" – rephrased in a slightly smoother way in the following ones, refer, more than to the issue of normative ethics, to the unattainable 'moral theodicy'.

6. *Sidgwick's missing criticism to Whewell*

Let us come now to Sidgwick's intellectual strategy, and let me try to locate Sidgwick's discussion of Whewell within such a strategy. Sidgwick insisted that in the *Methods* he had not been discussing intuitionism as a doctrine viewed at 'from outside', but was discussing instead the intuitional method as a method in which he himself could not avoid believing. He declares that "the general aim of the part of my treatise which deals with Intuitionism" is not

> criticising from the outside a particular school or sect of moral philosophers. My endeavour was rather to unfold a method of reaching practical decisions which I find (more or less implicit) in the ordinary thought of the society… *The doctrine which is called by the name Intuitionism is only one of those phases* (87).

Sidgwick's genuine criticism of Whewell's moral doctrine may be found instead in the *Outlines of the History of Moral Philosophy for English Readers*. Here he makes it clear that he believes the philosophical basis of intuitionism to have been worked out in full in the eighteenth century and that nothing important has been added after. He thinks that neither "Reid nor Stewart offers more than a very meagre and tentative contribution to that ethical science by which… the received rules of morality may be rationally deduced from self-evident first principles" (88) and that Whewell has been "more ambitious, but hardly more successful" (89), since his attempt "differs from that of this Scotch predecessors chiefly in a point where we may trace the influence of Kant – viz., in his rejection of

self-love as an independent rational and governing principle. And his consequent refusal to admit happiness, apart from duty, as a reasonable end for the individual" (90). It is true that Whewell has a "certain air of systematic completeness", and his five basic moral ideas try to depict a system of normative principles that be as complete as possible, but at a closer look

> we find that the principle of order, or obedience to government, is not seriously intended to imply the political absolutism… The formula of justice is given in the tautological or perfectly indefinite proposition "that every man ought to have his own"… however… this latter formula must be practically interpreted by positive law, though he inconsistently speaks as if it supplied a standard for judging laws to be right or wrong… Purity… merely particularises that supremacy of reason over sensuous impulses which is involved in the very notion of reasoned morality as applied to a being whose impulses are liable to deviate from rational duty (91).

Thus,

> if we ask for a clear and definite fundamental intuition, distinct from regard for happiness, we find really nothing in Whewell's doctrine except the single rule of veracity (including fidelity to promises); and even of this axiom the character becomes evanescent on closer inspection, since it is not maintained that the rule is practically unqualified, but only that it is practically undesirable to formulate its qualifications (92).

And so the general judgment Sidgwick passes on nineteenth century intuitionism is that

> the doctrine of the intuitional school, down to the middle of the present century, had been developed with less care and consistency than might have been expected, in its statements of the fundamental axioms or intuitively known premises of moral reasoning. And if the controversy which this school conducted with the utilitarianism of Paley and Bentham had turned principally on the determination of the matter of duty, there can be little doubt that it would have been forced into more serious and systematic effort to define precisely and completely the principles and method on which we are to reason deductively to practical conclusions. But in fact the difference between intuitionists and Utilitarians as

to the method of determining the particulars of the moral code was complicated with a more fundamental disagreement as to the very meaning of 'moral obligation' (93).

Let us examine now *The Methods* (94). On the one hand it is apparent that the book is not a utilitarian work. It was not so for theoretical reasons, namely that Sidgwick's own coherentist way of justifying the principle of utility, alternative to Mill's 'proof' and to Bentham's 'axiomatisation' of the principle itself was not – to Sidgwick's eyes – completely successful, at least as far as it worked against the intuitionist opponent but it did not work against the egoistic one. But it was so also for pragma-rhetoric reasons, namely because Sidgwick wanted to make the utilitarian doctrine, still perceived as a radical one, palatable to the audience by submerging its novelty under a heavy cloak of opinions supposedly supported by commonsense. As a result, and curiously enough, utilitarianism is criticized more in depth than intuitionism. Indeed Sidgwick works out a destructive criticism of the former doctrine heading to the conclusion that it lacks a real justification, is impossible to apply, and yet is the only way of talking about morality that makes any sense, since in order to make sense, an ethical theory should appraise actions on the basis of their consequence. The reader who would expect a parallel systematic discussion of intuitionism may be deceived in finding instead a discussion of "dogmatic intuitionism" that is exemplified by recourse to beliefs allegedly shared by the enlightened common opinion or by jurists, such as Blackstone who never had anything to do with the intuitionist philosophers This is strange enough, but Sidgwick had his own (more or less good) reasons for that. The fact is that a discussion of historically given intuitionism is never at issue here and Sidgwick does not take the pain to be fair to "Intuitional philosophers" because in this work he is considering their doctrines only occasionally and as examples of those procedures he believes to be practised by enlightened common sense and in whose (limited) validity he believes too. Thus the kind of intuitionism he discusses here is as a puppet he has tailored to his own purposes, not out of sheer bad faith, but instead as a kind of unintended effect of his own strategy vis-à-vis intuitionism conceived in terms of rescuing what is 'living' in intuitionism itself while discarding what is 'dead' (namely, undue philosophical overgrowth). Within the framework of such an approach to intuitionism, Whewell's doctrine seems to be not the real thing, but some kind of hybrid. In Sidgwick's view, it consists half of the naïve 'perceptive' intuitionism that is allegedly the 'doctrine' uncritically

adopted by common sense, the doctrine according to which good and bad actions are perceived immediately as such, and the other half consists in a philosophical theory, which in turn is useless in order to ground the doctrine.

In more detail, when Sidgwick wants to illustrate some philosophical doctrine defended by so-called "dogmatic intuitionism" he prefers to refer to Clarke as to the proponent of a more solid doctrine on the foundations of ethics and he refers to Reid as to the proponent of a more detailed reconstructions of the data of ethics such as may be reconstructed on the basis of common sense morality. Price and Whewell, strangely enough, are discussed less than Clarke and Reid, and are never presented as the proponents of a specific kind of intuitionism. The reason may be that – as I have already illustrated – the *old* intuitional school for Sidgwick included the Cambridge Platonists and allies, the *new* school included the Scottish common sense philosophers. Whewell, who was the avowed source of Sidgwick's dislike for "intuitional" doctrines, was neither discussed systematically nor given a consistent location either within the old school, to which his extreme rationalism seemed to draw him nearer, or with the new school, from which his own rationalism seemed to divide him. The reason may be that for both Price and Whewell common sense has a very limited importance, since their own kind of 'intuitionism' starts with the idea of self-evident rational propositions, not – unlike Reid, Stewart, Coleridge, Maurice, and perhaps Grote – with that of beliefs universally shared by humankind, and accordingly neither Price nor Whewell fits well Sidgwick's idea of an intuitionist.

What Sidgwick does is mentioning Whewell a number of times with reference to individual issues. In the seventh edition he mentions him explicitly eight times, and besides he clearly refers to some of his theses on a handful of occasions. Only twice the explicit mention is followed by a footnote with some precise reference. One of these, coming after mention of "cheerfulness, and the cultivation of the social affections" is apparently mistaken since it refers to Whewell's chapter where chastity is discussed, which is in fact the subject of Sidgwick's following paragraph (95); clearly enough, one more footnote referring to Kant's *Doctrine of Virtue* should show up at this point, since Sidgwick mentions the doctrine according to which appetite should be satisfied as a means of fostering "cheerfulness and the cultivation of the social affections", which is indeed Kant's doctrine (96). Let us examine six different topics about which Whewell's claims are discussed:

1. The first is the existence of a system of moral intuitions, which Sidgwick refuses while formulating the idea that common sense made consistent may be the best proxy for such system. He writes:

The orthodox moralists such as Whewell (then in vogue) said that there was a whole intelligible system of intuitions: but how were they to be learned? I could not accept Butler's view as to the sufficiency of a plain man's conscience: for it appeared to me that plain men agreed rather verbally than really.

In this state of mind I had to read Aristotle again… What he gave us there was the Common Sense Morality of Greece reduced to consistency by careful comparison (97).

2. On another occasion Whewell is mentioned as arguing the same as the Kantians, namely that a man "is a free agent in so far as he acts under the guidance of reason" (98), and as offering the justification that we ordinarily "consider our Reason as being ourselves rather than our desires and affections: we speak of Desire, Love, Anger, as mastering us, or of ourselves as controlling them. If we decide to prefer some remote and abstract good to immediate pleasures, or to conform to a rule which brings us present pain (which decision implies exercise of Reason). We more particularly consider such acts as our *own* acts" (99). Sidgwick admits that such statements win assent "from ordinary readers", since what Whewell describes is our usual way of considering reason. Yet, even though he does not object to this idea of freedom as denoting "voluntary actions in which the seductive solicitations of appetite or passion are successfully resisted", he sees a further problem that the Kantians as well as Whewell seem to overlook, that is, how to account for the very concept of responsibility, if one does not admit that an agent may be free to choose between acting rationally and acting irrationally. He adds: " We may say, if we like, that when we yield to passion, we become 'the slaves of our desires and appetites': but we must at the same time admit that our slavery is self-chosen" (100). Omitting discussion of the objection to Kant's view of freedom, which goes beyond the scope of the present essay, we may ask whether this is a fair objection to Whewell. In fact, Whewell adds:

If we ask why we thus identify ourselves with our rational part, rather than with our desires and affections; we reply, that it is because the Reason alone is capable of that reflex act by which we become conscious of ourselves. To have so much thought as to distinguish between

ourselves and our springs of action, is to be rational… It is by the Reason that we are conscious; and hence we place the seat of our consciousness in the Reason (101).

Whewell would object that acting under control of desire and affection uncontrolled by Reason means being – so to say – "passive, and merely acted on" (102), or, to be more precise, an agent in such a situation "is not really passive" but just adopts the suggestions of Desire or Affection, and rejects the control of Reason; he thus does not cease being aware "that there is a Rule, and that he is violating it" (103). In other words, he would say that passion is not irresistible and human action, qua human, has as its "essential condition" some amount of rationality. In the *Lectures* the point had been framed in terms of a distinction between dependent and independent schemes of morality; the latter are those

which would regulate human action by an internal principle or relation, as conscience or a moral faculty, or duty, or rectitude, or the superiority of the reason to desire… We maintain, with Plato, that reason has a natural and rightful authority over desire and affection; with Butler, that there is a difference of kind in our principles of action; with the general voice of mankind, that we must do what is right, at whatever cost of pain and loss (104).

Sidgwick goes on discussing the issue of Determinism and Free Will, which "is widely believed to be of great Ethical importance" (105) even if he is not sure it can be really settled.

3. A third point with regard to which Whewell is mentioned is a criticism to intuitionists of resorting actually to utilitarian considerations when trying to prove the necessity of moral rules.

This is a leading motif from the controversy between Mill and Whewell, echoing Bentham's main argument in favour of the principle of utility, namely that those who deny this principle in fact do affirm it in other words. Mill quotes Whewell while declaring that rules are necessary for the peace of society and that, without the satisfaction of some desires made possible by an ordered social life, "man's life is scarcely tolerable" (106), and he adds that here Whewell contradicts what he affirms elsewhere, since moral rules "are here spoken of as means to an end… This is utility – this is pleasure and pain. When real reasons are wanted, the repudiated happiness-principle is always the resource" (107). Mill goes on

widening the scope of his detection of the principle of utility in Whewell to the conclusion that

> Almost all the *generalia* of moral philosophy prefixed to the Elements are in like manner derived from utility. For example: that the desires, until subjected to general rules, bring mankind into conflict and opposition; but that, when general rules are established, the feelings which gather round these "are sources not of opposition, but of agreement"… This is Benthamism – even approaching to Fourierism (108).

He adds, as a further proof of the "hybrid character" of Whewell's theory, the remark that also his classification of virtues and duties "are in principle utilitarian. Though Dr. Whewell will not recognize the promotion of happiness as the ultimate principle, he deduces his secondary principles from it, and supports his propositions by utilitarian reasons as far as they will go" (109).

Was such criticism by Mill justified? I would remark that in ch. 3 Whewell is trying to reach in the beginning something that is for him like an intermediate halting-place, that is the proof that human life in society needs systems of rules, in order to try to prove that *there are* "such Moral Rules as we have spoken of" – which he supposes to be something *still in need of proof* – and says that in order to arrive at such rules "we must proceed by series of several steps" (110). He goes on then trying to show how human action is by its very nature, constituted through rule following, how the various rules are subordinate to each other, and how they presuppose a basic rule of human action (111). At this stage he believes he has proved not only that life in human society requires rules, but also that the constitution of human beings implies a set of rules which are self-evident in their basic contents and have an authority given by what we would now call 'internal' reasons. In other words, he believes he has proved on the one hand, that human society needs *some* set of rules, on the other that such sets may be of worse or better kinds and that there is an a priori way to the discovery of the essential structure of the justified set of rules, to be given flesh and bones then through a survey of detailed conditions of life and institutions existing in a given society. This is the reason why Whewell believes that morality depends on law as to the proof of the existence of a need for morality and as to the specification of a part of its actual contents, but that on the other hand, really existing systems of laws may be properly appreciated on the standard provided by

morality, which is something Mill always refused to admit was Whewell's point.

Sidgwick, apparently giving Mill's criticism for granted, bluntly states that

> even moralists (as Whewell) who are most strongly opposed to Utilitarianism have in attempting to exhibit the "necessity" of moral rules, been led to dwell on utilitarian considerations (112).

Even if the assertion is made without reporting precise statements by either Whewell or others, the passage Sidgwick has probably in mind is precisely ch. 3 in the *Elements*, book I, which was attacked by Mill, and he seems to assume that Mill's criticism was the final word. Besides he may have had his own historical reconstruction in mind according to which in the 17th century both intuitionism and utilitarianism were already there and both were in "friendly alliance" (113) fighting against the selfish system first proposed by Hobbes, and both approaches were seen as alternative ways of supporting the existing morality. It was only with Paley and Bentham that utilitarianism was first presented as method for determining conduct, which was to "overrule all traditional precepts and supersede all existing moral sentiments" (114). Sidgwick seems to add to Mill's argument that it was precisely because of such alliance that in the first phase no preoccupation arose with finding some 'pure' intuitionist way to "a philosophical basis of morality", since the real danger was then Hobbesian doctrine, and an opposition between utilitarianism and intuitionism was not on the agenda before Paley's and Bentham's time.

4. A point connected with the former is Whewell's allegedly inadequate account of justice, as far as for intuitionism – as Sidgwick understands it – the idea of justice should *translate* what common sense understands for justice into a more rigorous definition. Sidgwick declares that

> it is an assumption of the Intuitional method that the term 'justice' denotes a quality which it is ultimately desirable to realise in the conduct and social relations of men; and that a definition may be given of this which be accepted by all competent judges as presenting, in a clear and explicit form, what they have always meant by the term, though perhaps implicitly and vaguely (115).

On a careful examination of the data of common sense, it turns out yet that justice is "a kind of Equality" or better "Impartiality in the observance or enforcement of certain general rules allotting good and evil to individuals" (116) and that it includes the principles of reparation and those of conservative justice (compliance with contracts and laws and "normal" expectations) as well as of ideal justice, which in turns comprises conflicting ideals, namely the ideal of freedom and that of reward to desert (117).

In secondary literature this account has gone unnoticed as a matter of course. Yet, even if Sidgwick ascribes this account to common sense, or to common sense sifted by philosophical scrutiny, it would be naive to assume that the "Intuitional writers" were clearly on the same side with both 'common sense' and Sidgwick. In fact, Whewell's account is somewhat different and it would be interesting to know whether Sidgwick had any specific objection to such an account. Whewell had defined justice as "the Desire that each person should have his own" (118), and the corresponding part of Whewell's Supreme Rule which belongs to this Virtue declares that "*each man is to have his own*" (119); more substantive contents of such virtue and of the corresponding Rule, that is, a specification of the rights everyone may claim in matters of property, derive from historically given institutions of each particular society, which vary according to the previous historical circumstances and the present conditions of life, but which are not to be taken as something given once for ever but instead are to be modified with a view at a closer approximation to the ideal of equality between human beings (120).

5. Another crucial topic is truthfulness and promises. On these issues I have already reminded that Sidgwick's criticism is that Whewell is unable to provide any content that would not turn out "evanescent at a more accurate examination". A duty to keep promises – Sidgwick acknowledges – is admitted by everybody; the obligation seems to unreflective common sense to be intuitively independent and certain; on the other hand, yet, existence of a number of exceptions seems to be commonly accepted: namely, when a promise contrasts with another obligation; when what has been promised is immoral; when circumstances have been modified; when the promise has been obtained by a lie etc. Common sense yet (note, common sense, not "intuitional" philosophy) seems to be unable to reach a consensus on what are precisely the cases where a promise must be kept, and "if one of these conditions vanishes it seems that consensus becomes

evanescent and that common moral intuitions of reflective persons become obscure and diverge" (121).

Among these a typical one is that of a promise formulated without prior knowledge of relevant facts or before important elements of the context were modified. On this case, according to Sidgwick, common sense "seems to give no clear answer" (122). Why, if common sense fails, also other tentative ways of giving an answer should be dropped is never spelled out. That is, the "Intuitional" moralists did try to provide answers, and these were based on rational a priori arguments, not on surveys of what common sense seems to suggest. Here Sidgwick's ambivalence plays a decisive role. He was sure that it was so because the intuitive method was *his own* method as well as that *of the most educated part* of society, and accordingly he only needed to ask himself and consult his acquaintances over a cup of tea, a method not different from the one Hume himself used in order to discover what *belief* normally yields. But in Sidgwick's case conclusions reached through his own kind of amateur sociological survey concerning intuitions reached by common sense were then applied with no further step to the claims of intuitional philosophers as well, without apparently any strict duty to read what they had really said on the issue. This is particularly striking with reference to this issue and to Whewell, since in ch. 15 of book II of the *Elements* he had claimed to have given a solution precisely to this problem by his "Principle of Truth", namely by establishing that in such cases as the one of the promise obtained by giving false information, any duty arising from the promise itself should always be understood as conditional duty, bound to truthfulness of the conditions made known at the time the promise is formulated. Whewell argues his conclusion also referring to cases widely discussed in the casuistic literature and already found in Cicero (123). In this way Sidgwick gives an answer to one question and makes the reader believe he has answered also a different one, namely he gives the impression he does not need to criticize Whewell's solution, which is something different from what common sense suggests or fails to suggest. Whewell indeed never claimed that common sense *has already settled* the issue, but *only that it be possible to solve it* by means of distinctions that are rationally justifiable but also highly abstract and novel, and no way already familiar to common sense. The fact that what he argues for is a novelty for common sense does not imply the validity of Mill's accusation of falling back into considerations of expediency, since his argument is based solely on criteria of inner consistency, and Sidgwick, were he to prove that the intuitionists are wrong on the point, should have carried out a criticism of

the arguments of Whewell, the casuists, Cicero, not of the opinions allegedly shared by common sense (124). It is surprising that on the one hand Sidgwick discards what intuitionism has to offer in order to settle the issue on the argument that it *seems* that common sense have nothing to say and on the other avoids criticizing in detail the solution proposed by the last proponent of this doctrine. Sidgwick – as I have already suggested – had his own justification for this, namely that in the *Methods* he did not want to examine intuitionism as a doctrine from outside, but as a "method" practised by common sense and to a certain extent accepted by Sidgwick himself. But here Sidgwick's ambivalence to intuitionism (sometimes a mistaken philosophical doctrine and sometimes a plausible albeit limited "method" for formulating moral judgements) becomes an unconscious excuse for dodging the main objections that could withstand his will to be a Millian, or cast a doubt on his conclusion that utilitarianism is a rickety building, and yet is the only roof left under which we may find shelter.

A closely related issue is establishing limits or exceptions to the duty of truthfulness. The rule 'to speak the truth' would not be difficult to apply, yet, even if "many moralists have regarded this, from its simplicity and definiteness, as a quite unexceptionable instance of an ethical axiom" (125).

Nonetheless, "reflection" shows that truthfulness cannot be raised to the status of a "definite moral axiom" (126) because common sense seems to admit that the right to truthfulness may be suspended under certain circumstances, such as those under which most of us "would not hesitate to speak falsely to an invalid" (127), and that we cannot establish "how we can decide when and how far it is admissible, except by considerations of expediency" (128). As a conclusion the rule of Veracity cannot be elevated into a "definite moral axiom" for there is no agreement as to when absolute sincerity is required (129). Also the Kantian argument of the self-destroying character of the rule of lying under certain circumstances is discarded claiming – in a perfectly Millian spirit – that it is no more than a "strong – but not formally conclusive – utilitarian ground for speaking the truth" (130).

Concerning Veracity Sidgwick makes a precise reference to Whewell. He writes:

> it is not uncommonly said that in defence of a secret we may not indeed *lie* (fn.: Whewell, *Elements*, book II, ch. xv, par. 299), i.e. produce directly beliefs contrary to fact; but we may "turn a question aside"… or

"throw the inquirer on a wrong scent"… These two methods of concealment are known respectively as *suppressio veri* and *suggestio falsi*, and many think them legitimate under certain circumstances: while others say that if deception is to be practised at all, it is mere formalism to object to any one mode of effecting it more than another (131).

Sidgwick's own opinion has already been presented, that is, he endorses the latter opinion. But let me remind, before trying to assess the goodness of Sidgwick's criticism, what Whewell had actually written. This is:

(i) that "the necessary conditions of a Rule of Human Action is the existence of a Common Understanding among men, such that they can depend upon each other's premeditated and predetermined actions" (132);

(ii) that the idea of Truth as a Virtue, which may also be named integrity or Truthfulness, is the idea of a conformity of "*our language to the universal understanding among men which the use of language* implies" (133);

(iii) that some kind of implicit contract that binds human beings to telling the truth, "the universal understanding among men which the use of language implies", is tacitly signed when they start using language, and a "Right to know the Truth is conveyed, by every speaker, to the person to whom he addresses the assertions" (134);

(iv) that lying, no less than not keeping a promise, is "a violation of the general understanding among mankind, which the use of language implies" (135);

(v) that lying always carries a moral stain on the liar, with an exception to be made for cases of necessity, as when it is made to save one's life, which is looked upon "as at least excusable, and allowable", or to save a friend from some great misfortune, which meets "with a more decided approval" (136);

(vi) that in cases of necessity which are also conflicts of duties, as far as a moral rule is transgressed not with a view at one's preservation, but "in order to preserve some other person from great impending evil" it is better for the moralist to abstain from laying down definite rules of decision, for doing so

would have an immoral tendency. For such a procedure would necessarily seem to make light of the Duties which were thus, in a general manner, postponed to other Duties; and would tend to remove the com-

punction, which any Moral Rule violated, ought to occasion to the Actor (137).

It may be worth noting that with regard to lying Sidgwick quotes Whewell in a precise way, but also that he mentions his conclusion as one of these opinions which moralists allegedly share with common sense, and that he does not discuss in any detail Whewell's solution to the conundrum raised by cases of necessity. Also here he has his own reason for not doing that, since he believes that a critique of the intuitionist moralists' arguments goes beyond his own self-appointed task, which is amending and systematizing the opinions shared by common sense. This could be a convincing enough reason, if only Sidgwick after that did not announce the conclusion that as a consequence "dogmatic intuitionism" – which, according to Sidgwick himself, is tantamount to the doctrines of the British rationalist and/or common-sense moralists from Clarke, Butler, Price to Reid and Whewell – does not stand up.

6. The relationship of morality and law. Sidgwick's criticism to Whewell's Idea of Order on which the latter grounded the claim that obedience to law be on principle an unconditional duty (138) echoes heavily the Mill-Whewell controversy. I recalled above how Mill had bluntly accused Whewell of implicitly defending slave owners and besides of heading to a vicious circle. Sidgwick denies that it be possible to settle conflicts between civil and moral law unless we have recourse to the utilitarian method since common sense only manifest a rather vague general consensus on the idea that law as such should be obeyed. A proof of such impossibility to reach precise shared conclusion is that "jurists" (note again, jurists, not intuitional moralists) "have contrasting opinions as to the fact whether we are strictly bound to obedience to laws when they command what is not otherwise a duty or forbid what is not otherwise a sin" (139). On the basis of "so much difference of opinions" Sidgwick announces that

It seems idle to maintain that there is any clear and precise axiom or first principle of Order, intuitively seen to be true by the common reason and conscience of mankind. There is, no doubt, a vague general habit of obedience to laws as such… but when we try to state any explicit principle corresponding to this general habit, the *consensus* seems to abandon us (140).

Note that the "axiom or first principle of Order" mentioned is not yielded by some mental experiment enacted by Sidgwick but is a notorious Whewellian doctrine that contemporary readers could easily have associated with its author's name. The principle is defined by the latter as

a disposition to conform, both to positive human Laws… and to special Moral Rules, as the expression of the Supreme Rule… And the corresponding part of the Supreme Rule is: *We must accept positive Laws as the necessary conditions of Morality* (141).

The remark is not out of place that the principle had *not* been introduced by Whewell as a means of settling the issue under discussion by Sidgwick. The latter was a doubt concerning the subsistence of an obligation to obey the civil law in a number of cases, a familiar problem in casuistry, to which the casuists had given more complex answers than those ascribed by Sidgwick to the jurists, while mentioning just Austin, Hobbes, and Blackstone (142). Whewell's aim was instead to examine a more general issue. Whewell had added:

We must conform our Dispositions to the Laws; obey the Laws cordially, or administer them carefully, according to the position we may happen to hold in the community. This disposition may be denoted by the term Order, understood in a large and comprehensive sense. But further: not only positive human Laws, but subordinate moral Rules, are necessary conditions of morality. We cannot conform our actions, intentions, desires to the Supreme Rule, without having in our thoughts subordinate Rules, which are partial expressions of the Supreme Rule; and to such subordinate Rules, it is our Duty to conform our Intentions and Desires. The disposition to do this may also be included in the term Order, taken in its largest sense (143).

That is, what Whewell was concerned with in the quoted passage where he introduces the "axiom or first principle of Order" was the relationship of general and particular laws, be they *civil* or *moral* laws. Sidgwick seems to be ignorant of the circumstance that the problem mentioned had been treated by Whewell elsewhere, namely in the *Elements*, book IV, chapter 1. The fourth book of the 3^{d} edition is something new, that was absent in the text Mill read – or did not read in full – but Sidgwick may have been just following Mill blindly without noticing that he had a detailed answer to Mill's criticism by Whewell at hand. Mill had

made it a point of honour to declare that he had confined himself to those pages of the *Elements* which could be evidence on one point, that is, how he "argues in condemnation of any external standard, and especially of utility, or tendency to happiness, as the principle or test of morality [as well as] how he fares in his attempt to construct a coherent theory of morals on any other basis" (144). On a close reading, Mill appears to refer only to few pages in book I and he never mentions book IV, chapter 21 where the issue of the relationship of law and morality is discussed in more detail than in the couple of pages from the "Preliminary Lecture" in the *Lectures on the History of Moral Philosophy in England* to which Mill, strangely enough, limits himself when discussing this point (145).

In this chapter Whewell had done what he did later in a more detailed way in book IV of the 3d edition, that is, he had illustrated how morality depends on law for one aspect, the definition of rights, which are indeed the subject-matter of moral rules, but not for a different aspect, namely in so far as morality provides a standard on which historically given laws may be appraised and with a view at converging with which we may wish that they be modified, and "thus, for the moment, at any time, Morality depends upon Law; but in the long run, Law must be regulated by Morality" (146).

It is important to remark that – contrary to what Sidgwick seems to believe – for Whewell the standard on which positive law is to be judged is not the axiom of Order, which is called to carry out rather the function that I have illustrated, but that of Justice (147).

Besides, it may be mentioned that Whewell believed he had settled the problem whose existence Sidgwick denounces with regard to justice, namely that when we try to make the apparent principles in which it seems to consist more precise, "we find ourselves involved in grave difficulties" (148). In Book II ch. 21 Whewell responds to the ancient objection according to which the law of nature, being positive laws different in different states, exists nowhere. The answer is that the trouble may be dissolved in the light of the general claim of circularity between Idea and Fact. In more detail the answer is that

> the *Conceptions* of the Fundamental Rights, which Law establishes, are necessary and universal for all men; but that the *Definitions* of these Rights are Facts, which grow out of the History of each community, and may be different in different times and places (149).

Again, Sidgwick does not seem to be aware of Whewell's attempt at solving the problem, that is, at reconciling variability and universality, and thus never tries to criticize the solution offered. As I have argued, also on this issue, he believes he need not criticize intuitionist doctrines since he is not really interested in such doctrines but believes instead he should draw on common usage and try to reach a definition which be acceptable to all "competent judges", and in doing so, so to say, "clip the ragged edge of common usage, but we must not make excision of any considerable portion" (150). So much is what he thinks being required by the "Intuitional method" (151), but this, once again, is *his own* method, not the method of the intuitionist moralists.

7. *Concluding remarks*

To sum up, my claims were the following:
1. Sidgwick's notion of "dogmatic intuitionism", an expression reverently repeated by commentators, is a queer notion; it is the result of one of those divisions of one into two that philosophers use to stage every time they want to keep an old doctrine while claiming originality; in this case, *dogmatic* intuitionism was the 'bad company' to which all that Sidgwick did not like of "the intuitional school of morality" should be entrusted, to be distinguished from the 'good company', *philosophical* intuitionism that was to take over all that Sidgwick liked of this school; that is, it was a way of disguising the fact that Sidgwick's final doctrine was ethical intuitionism.

2. Sidgwick's reconstruction of the history of 'intuitionist' doctrines is an odd one in so far as he distinguishes between an earlier more philosophical school and a later school more based on common sense; it is clearly Reid and Dugald Stewart that he has in mind, and Whewell, with his bold apriorism, seems to drop out of the picture; besides he ignores Price totally.

3. Sidgwick has recourse to a strange enough argument for justifying his lack of a real criticism of intuitionist doctrines; that is, he is interested in assessing the role of intuitions in common sense, not the role assigned to them by philosophers, but then he constantly shifts from allegedly proved conclusions concerning the limits of common sense to unwarranted conclusions on the intuitionists' mistakes.

4. Whewell had his own version of a rationalist (not common sense) intuitionist ethics, which followed Price closely and also incorporated a

few Kantian insights; on this version, moral dilemmas could on principles be settled through argument, but untutored common sense did not possess already a clear solution to such dilemmas.

5. Such version incorporated solutions or alleged solutions to a few of those difficulties of moral reasoning that Sidgwick believed were decisive in proving the inability of both common sense and intuitionist ethics in settling moral dilemmas; Alan Donagan has claimed that Sidgwick's polemic is vitiated by the fact of ascribing to Whewell a claim that the latter had never advanced, that is, that common-sense morality may afford a solution to moral dilemmas; what Whewell did is proposing a way of dissolving, on the basis of intuitionist procedures (not of common sense morality), that is, starting with clauses and limitations to duties that may be logically derived from the general formulation of general precepts, most of apparent moral dilemmas; on examples such as the duty to keep promises extracted through reticence concerning relevant information Whewell's answer is that such a promise is not a real promise since full knowledge of relevant facts is one of the conditions of the act of promising; Sidgwick does not discuss Whewell's argument, and indeed it is impossible to prove that he ever read the relevant chapter, and limits himself to noting that common sense lacks an answer, but if intuitionists were right, common sense should already know the right answer (which is not what Whewell claimed).

6. Sidgwick is far from immune to rhetoric, and indeed his work is a powerful experiment in persuasion, adopting as a systematic strategy the stratagem of introducing subversive ideas – among other things, concerning current standards of sexual morality – hidden under a heavy burden of received opinions and justified repeatedly by appeal to one's faithfulness to the duties carried by the status of philosopher or scientist; also the choice of writing dry and as-boring-as-possible treatises is a rhetorical trick no less than any declamation about the beauty of virtue; the message conveyed is: "I am not a preacher, I am a scientist"; that is, Sidgwick's trump is one of the rhetorical stratagems recommended by Schopenhauer: if you lack specific objections, shift from the point under discussion to general considerations on the limits of human knowledge, suggesting by implication that your opponent's claim cannot have strong reasons on its side, since nobody's claim does.

7. Sidgwick wanted basically to defend Millian ideals, and believed his own theoretical work to be also a powerful exercise in persuasion; in order to do that, he believed he had to sacrifice all of Mill's (as well as Bentham's) strictly philosophical ideas on ethics, adopting instead Whewel-

lian intuitionism as a "philosophy of morality" (i.e. metaethics); but he used such philosophy of morality in order to support conclusions in morality (i.e. normative ethics) opposite to Whewell's and close to Mill's; by doing so, he wanted to defend a familiar view of ethics, shared by both Utilitarians and Intuitionists, against new approaches, Spencer's evolutionism and Bradley's idealism, as well as – had it been possible – against an old/new approach, ethical egoism, against which he confessed his weapons were blunt; yet, he was keen in giving the impression that his newly assembled machine as Utilitarianism-on-a-new-basis, not as Intuitionism-improved; on the main philosophical issue, the need for intuitions in ethics, he acknowledged the victory of Whewell on Mill but – what t is typical of all philosophical controversies – he condemned the sinner while condoning the sin, and appropriated Whewell's ideas while declaring the latter to be a shallow thinker. "We buy our opinions wholesale" was one of the famous remarks by Montaigne; right or wrong, this is precisely what Sidgwick did, since at some point he came back to Whewell on all that mattered for a philosopher, but he remained all his life aligned with the Utilitarian camp on real-world issues. The result was leaving to twentieth-century analytic ethics a legacy of intuitionist *ideas* combined with utilitarian *opinions* and sanctifying Mill's image as a discoverer of truths he would never had been able to discover by himself (i.e. without his controversy with Whewell), and damning Whewell's figure to oblivion as but the efficient cause of "one of the thousand waves on the dead sea of commonplace" (152).

Acknowledgements: I had the opportunity to take advantage, besides the guest editor's Gianfranco Pellegrino competent revision, also of useful comments by Massimo Reichlin.

(1) A. Donagan, *Whewell's Elements of Morality*, "Journal of Philosophy" 71\1 (1974): 724-736, p. 734.
(2) H. Sidgwick, *The Methods of Ethics*, 7th ed. (1907), in *Works,* 15 vols. (Bristol: Thoemmes Press, 1996, vol. I), p. xiii; unless otherwise stated, reference will be this edition.
(3) B. Schultz, *Henry Sidgwick. Eye of the Universe* (Cambridge: Cambridge University Press, 2004), particularly pp. 181, 142-146, 155.
(4) *Ibidem*, p. 511. For an example of a study of such pragma-rhetoric structure of texts as the one Schultz proposes to carry out for Sidgwick,

see S. Cremaschi and M. Dascal, *Malthus and Ricardo on Economic Methodology*, "History of Political Economy" 28\3 (1996): 475-511; Idd., *Persuasion and Argument in the Malthus-Ricardo Correspondence*, in W.J. Samuels and J.E. Biddle (eds.), *Research in the History of Economic Thought and Methodology*, Stanford, Co: JAI Press, vol. 16, 1998, pp. 1-63; Idd., *Malthus and Ricardo: Two Styles for Economic Theory*, "Science in Context" 11\2 (1998): 229-254; M. Dascal and S. Cremaschi, *The Malthus-Ricardo Correspondence: Sequential structure, argumentative patterns, and rationality*, "Journal of Pragmatics" 31 (1999): 1129-1172.

(5) A. Donagan, *Whewell's Elements of Morality*; Id., *Sidgwick and Whewellian Intuitionism*, in B. Schulz (ed.), *Essays on Henry Sidgwick* (Cambridge: Cambridge University Press), pp. 123-142.

(6) J. B. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy* (Oxford: Clarendon, 1977).

(7) B. Schulz, *Henry Sidgwick*, p. 143.

(8) H. Sidgwick, *The Methods*, p. xv.

(9) W. Whewell, Preface to J. Mackintosh, *Dissertation on the Progress of Ethical Philosophy, chiefly during the Seventeenth and Eighteenth Centuries* (1836, Bristol: Thoemmes, 1991); Id., *Elements of Morality, including Polity*, 3d edition (London: Parker 1854); unless otherwise stated reference will be this edition; numbers of paragraphs in the 4th edition remain unchanged.

(10) A. Sedgwick, *Discourse on the Studies of the University* (1833), ed. by E. Ashby and M. Anderson (Leicester: Leicester University Press 1969).

(11) See J. Brown, *On the Motives to Virtue, and the Necessity of Religious Principle* (1751), in J. Crimmins (ed.), *Utilitarianism and Religion* (Bristol: Thoemmes 1998), pp. 55-104; J. Gay, *Concerning the Fundamental Principles of Virtue or Morality* (1731), *Ibidem*, pp. 33-48; W. Paley, *The Principles of Moral and Political Philosophy*, 1785, ed. by D.L. LeMahieu (Liberty Fund, Indianapolis, IN, 2002).

(12) See L.J. Snyder, Whewell, William, *Stanford Encyclopedia of Philosophy* (2000, rev. 2006) (http://plato.stanford.edu/entries/Whewell/).

(13) W. Whewell, *Two Lectures to two Courses of Lectures on Moral Philosophy*, in Id., *Collected Works*, 16 vols., ed. by R. Yeo (Bristol-London: Thoemmes - Routledge, 2001), vol. XI, p. 28.

(14) W. Whewell, *Elements of Morality*, art. 97.

(15) *Ibidem*.

(16) W. Whewell, *Lectures on the History of Moral Philosophy in England* (1852, Bristol: Thoemmes, 1990), p. 223.

(17) *Ibidem*, p. 221.

(18) *Ibidem*, p. 222.
(19) *Ibidem*.
(20) W. Whewell, *Elements, including Polity*, 1st ed. (1845), 2 vols. (New York: Harper, 1861, vol. I, pp. 7-8); the idea that a system of normative ethics should be completed in a consistent way as a condition for starting discussion of metaethical issues is suggested, while referring to Noam Chomsky instead of Wheweel, in J. Rawls, *A Theory of Justice* (Cambridge, MASS: Harvard University Press, 1971), pp. 46-48
(21) H. Sidgwick, *The Methods*, p. v.
(22) This is why Rawls made this point without apparent awareness of having being 'forerun' by Whewell. On the aspects under which first Whewell discovered a coherentist approach and then Sidgwick adopted it, while trying to use it in order to reach conclusions opposite to Whewell's, see J.B. Schneewind, *First Principles and Common Sense Morality in Sidgwick's Ethics*, "Archiv für Geschichte der Philosophie", 45\2 (1963): 137-156; Id., "Whewell's Ethics", in *Studies in Moral Philosophy*, "American Philosophical Quarterly. Monograph Series", 1 (1968): 108-141. Note that at the time Schneewind was writing a similar kind of approach was still waiting to be rediscovered by Rawls.
(23)W. Whewell, *Elements*, art. 66.
(2) *Ibidem*, art 62.
(25) *Ibidem*.
(26) *Ibidem*, art. 63.
(27) *Ibidem*, art. 128.
(28) *Ibidem*, art. 75.
(29) *Ibidem*, art. 76.
(30) *Ibidem*, art. 68-70.
(31) *Ibidem*, p. 1.
32) *Ibidem*, art. 322.
(33) *Ibidem*.
(34) *Ibidem*, art. 323.
(35) *Ibidem*.
(36) *Ibidem*, art. 326.
(37) *Ibidem*, art. 327.
(38) *Ibidem*, art. 279.
(39) *Ibidem,* art. 201, 296-7.
(40) *Ibidem*, art. 281.
(41) *Ibidem*.
(42) *Ibidem*, art. 294.
(43) *Ibidem*, art. 299.

(44) *Ibidem*, art. 303.
(45) W. Paley, *The principles of moral and political philosophy* (1785), ed. by D.L. LeMahieu, (Indianapolis, IN: Liberty Fund, 2002), book III, ch. 5, sect. 3.
(46) W. Whewell, *Elements*, , art. 308.
(47) *Ibidem,* art. 489.
(48) *Ibidem,* art. 390.
(49) See E.W. Strong, *William Whewell and John Stuart Mill: Their Controversy over Scientific Knowledge*, "Journal of the History of Ideas" 16 (1955): 209-31; G. Buchdahl, "Deductivist versus Inductivist Approaches in the Philosophy of Science as Illustrated by Some Controversies Between Whewell and Mill," in M. Fisch, S. Schaffer (eds.), *William Whewell: A Composite Portrait* (Oxford: Oxford University Press, 1991), pp. 311-44; J. Losee, *Whewell and Mill on the Relation between Science and Philosophy of Science*, "Studies in History and Philosophy of Science" 14 (1983): 113-26; L.J. Snyder, *Reforming Philosophy: A Victorian Debate on Science and Society* (Chicago: University of Chicago Press, 2006), chs. 1-3.
(50) W. Whewell, *Lectures on the History of Moral Philosophy*, pp. 202-265; the rest of the present section draws on materials from S. Cremaschi, *The Mill-Whewell Controversy on Ethics and its Bequest to Analytic Philosophy*, in E. Baccarini and S. Prijić Samaržja (eds.), *Rationality in Belief and Action*, University of Rijeka, Rijeka: Faculty of Arts and Sciences - Croatian Society for Analytic Philosophy, 2006, pp. 45-62.
(51) J.S. Mill, *Autobiography*, in *Collected Works of John Stuart Mill*, ed. by J.M. Robson *et al.* (Toronto: University of Toronto Press, 1967-, vol. I, pp. 1- 290), p. 209, emphasis added.
(52) Discussion of Benthamite ethics between 1788 and the half of the nineteenth century was rather limited. F. Jeffrey in *Bentham's Traité de Législation civile et pénale,* "Edinburgh Review" 4 (1804), n. 7: 1-20, had reviewed Bentham's treatise of legislation, but ethics was not a major concern; a discussion of Bentham's political ideas is in J. Mackintosh, *Bentham's Plan of a Parliamentary Reform*, "The Edinburgh Review" 31 (1818), n. 61: 165-203; the first extended discussion of Benthamite ethics is in a series of three articles: T.B. Macaulay, *Mill's Essay on Government: Utilitarian Logic and Politics*, "The Edinburgh Review" 49 (March 1829), n. 97; Id., *Bentham's Defence of Mill: Utilitarian System of Philosophy*, "The Edinburgh Review" 49 (June 1829), n. 98; Id., *Utilitarian Theory of Government , and the 'Greatest Happiness' Principle*, "The Edinburgh Review" 49 (October 1829), n. 99; they are all reprinted in J. Lively and J. Rees eds., *Utilitarian Logic and Politics*, Oxford: Clarendon Press 1979,

pp. 97-129, 151-178, 193-223; note that the controversy started with criticism by MacIntosh of Benthamite political doctrines and only with the last of his articles shifted to ethics; after that, Bentham's ethics was first made the subject of extended treatment in an historical overview of British philosophy by J. Mackintosh, *Dissertation on the Progress of Ethical Philosophy, chiefly during the Seventeenth and Eighteenth Centuries* (1836, Bristol: Thoemmes, 1991), pp. 284-313; on the above mentioned literature see S. Cremaschi, *Utilitarianism and its Nineteenth-Century Critics*, "Notizie di Politeia" 24 (2008), n. 90: 31-49.

(53) J.S. Mill, *Autobiography*, p. 232.

(54) J.S. Mill, *Sedgwick's Discourse* (1835), in *Collected Works*, vol. X, pp. 31-74, p. 51.

(55) See W. Whewell, *On the Foundations of Morals. Four Sermons* (1837), in Id., *Collected Works*, vol. XI, p. vii; Id., *Elements*, vol. II, p. 303.

(55) W. Whewell, *Lectures on the History of Moral Philosophy*, p. 215.

(56) See *ibidem*, pp. 210-212.

(58) W. Whewell, *Elements,* p. 12.

(59) W. Whewell, *Lectures on the History of Moral Philosophy*, p. 215.

(60) See *ibidem*, p. 216.

(61) J.S. Mill, *Whewell on Moral Philosophy* (1852) in *Collected works of John Stuart Mill*, vol. X, pp. 167-201, p. 169.

(62) See *ibidem*, pp. 169-170.

(63) See *ibidem*, p. 171.

(64) *Ibidem*, p. 192

(65) *Ibidem*.

(66) W. Whewell, *Elements,* vol. II, pp. 300-303.

(67) *Ibidem*.

(68) *Ibidem*, p. 305.

(69) L.G. Snyder, "Whewell, William", p. 10; as already noted, the discovery of the coherentist approach in Whewell has been made in J.B. Schneewind, "Whewell's Ethics", in *Studies in Moral Philosophy*, "American Philosophical Quarterly. Monograph Series", 1 (1968): 108-141; cf. *Sidgwick's Ethics and Victorian Moral Philosophy*, pp. 89-121.

(70) See L.G. Snyder, *Reforming Philosophy*, p. 266.

(71) S. Collini, *Public Moralists. Political Thought and Intellectual Life in Britain 1850-1930* (Oxford: Clarendon Press, 1991), p 122; for a remarkable overview of Mill's role as the first of the Victorian public moralists see pp. 121- 169.

(72) See L.G. Snyder, *Reforming Philosophy*, ch. 4.

(73) This reference to *Nichomachean Ethic*s as an example to be followed in so far as it was a successful attempt at reducing the Common Sense Morality of Greece to "consistency" by "careful comparison" is in the Preface to the sixth edition, in H. Sidgwick, *Methods of Ethics*, p. xix.
(74) *Ibidem*., p. xv.
(75) *Ibidem*.
(76) *Ibidem*, p. 103.
(77) *Ibidem*.
(78) A. Donagan, *Whewell's Elements*, pp. 734-735.
(79) W. Whewell, *Two Introductory Lectures to two Courses of Lectures on Moral Philosophy*, pp. 43-44.
(80) H. Sidgwick, "Letter to H.G. Dakyns", Aug. 24, 1861, in . Sidgwick, E. M. Sidgwick (eds.), *Henry Sidgwick: A Memoir* (1906; Bristol: Thoemmes,1906), p. 68.
(81) Letter to H.G. Dakyns, March 1862, *ibidem*, p. 75.
(82) Letter to H.G. Dakyns, December 1862, *ibidem*, p. 90.
(83) J.M. Keynes, *Essays in Biography* (*The collected Writings*, vol x; London: MacMillan, 1972), p. 173.
(84) See S. Cremaschi, *Utilitarianism and its British Nineteenth-Century Critics*, pp. 38-39 and 42-43; J.B. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy*, pp. 89-121; Schulz, Henry Sidgwick, pp. 45-54
(85) A. Sidgwick, E. M. Sidgwick (eds.), *Henry Sidgwick: A Memoir*, p. 472.
(86) *Ibidem*.
(87) H. Sidgwick, *Professor Calderwood on Intuitionism in Morals* (1876), in *Collected Works*, vol. XIV, p. 563; emphasis added.
(88) H. Sidgwick, *Outlines of the History of Ethics for English Readers* (1886; Bristol: Thoemmes, 1996), p. 233.
(89) *Ibidem*.
(90) *Ibidem*.
(91) *Ibidem*, p. 234.
(92) *Ibidem*.
(93) *Ibidem*.
(94) The rest of this section is an expanded version of sect. 7 in S. Cremaschi, *Sidgwick e il progetto di un'etica scientifica*, "Etica e Politica/Ethics & Politics" 7/1 (2006), pp. 1-36.
(95) *Ibidem*, p. 329; cf. Whewell, *Elements*, book II, ch. X: "Duties connected with purity".
(96) H. Sidgwick, *The Methods of Ethics*, p. 58.

(97) W. Whewell, *Elements*, art. 63. Sidgwick adds that "it is also true – as I afterwards say – that we sometimes identify ourselves with passion or appetite in conscious conflict with reason: and then the rule of reason is apt to appear an external constraint, and obedience to it a servitude, if not a slavery" (*The Methods*, p. 58 fn); in the first edition (p. 44) the comment in the footnote was absent; in its place there was a more extended discussion in the text of the case of "many persons, to whom, from a preponderance of the emotional and active elements in their nature, the state of reflection in which action is most deliberate is essentially irksome and depressing" (pp. 44-45).
(98) *Ibidem*, p. 59.
(99) W. Whewell, *Elements*, art. 63.
(100), H. Sidgwick, *The Methods*, p. 59
(101) W. Whewell, *Elements*,, art. 64
(102) *Ibidem*.
(103) *Ibidem*.
(104) W. Whewell, *Lectures on the History of Moral Philosophy*, p. x.
(105) H. Sidgwick, *The Methods of Ethics*, p. 58; in the 1st edition Sidgwick was much clearer in declaring that he believed that ethics could go without any solution of the dispute on free will (p. 45) and that there seems to be "no general connexion between systematic ethics and the disputed question of Free Will" (p. 57).
(106) W. Whewell, *Elements*, 1st ed., art. 65; cf. 3d ed., art. 66.
(107) J.S. Mill, *Whewell on Moral Philosophy*, p. 192.
(108) *Ibidem*, pp., 193-3.
(109) *Ibidem*, p. 193.
(110) W. Whewell, *Elements*, 1st ed., book I, ch. 3, art. 69.
(111) *Ibidem*, art. 70-77.
(112) H. Sidgwick, *The Methods*, p. 86.
(113) *Ibidem*.
(114) *Ibidem*.
(115) *Ibidem*, p. 263.
(116) *Ibidem*, p. 293.
(117) See *Ibidem*, pp. 293-294.
(118) W. Whewell, *Elements*, art. 119.
(119) *Ibidem*.
(120) *Ibidem*, art. 386 and 397
(121) H. Sidgwick, *The Methods*, p. 311.
(122) *Ibidem*, p. 308.
(123) W. Whewell, *Elements*, ch. 15.

(124) This point was argued forcefully in A. Donagan, *Whewell's Elements*, pp. 734-735.
(125) H. Sidgwick, *The Methods*, p. 315.
(126) *Ibidem*.
(127) *Ibidem*, p. 316.
(128) *Ibidem*.
(129) *Ibidem*, p. 317.
(130) *Ibidem*, p. 319.
(131) *Ibidem*, p. 317.
(132) W. Whewell, *Elements*, art. 216.
(133) *Ibidem*, art. 296.
(134) *Ibidem*, art. 301.
(135) *Ibidem*.
(136) *Ibidem*, art. 323.
(137) *Ibidem*.
(138) H. Sidgwick, *The Methods*, pp. 295-303.
(139) H. Sidgwick, *The Methods*, p. 302.
(140) *Ibidem*, p. 303.
(141) W. Whewell, *Elements*, art. 122.
(142) H. Sidgwick, *The Methods*, pp. 300-302.
(143) W. Whewell, *Elements*, art. 122.
(144) J.S. Mill, *Whewell on Moral Philosophy*, pp. 169 and 191.
(145) *Ibidem*, pp. 188-189.
(146) W. Whewell, *Elements*, art. 648; cf. 1st edition, art. 217-222.
(147) W. Whewell, *Elements*, art. 119.
(148) H. Sidgwick, *The Methods*, p. 294.
(149) W. Whewell, *Elements*, art. 382.
(150) H. Sidgwick, *The Methods*, pp. 264.
(151) *Ibidem*.
(152) J.S. Mill, *Whewell on Moral Philosophy*, p. 169.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 185-209

Sidgwick's Philosophical Intuitions

Anthony Skelton
Department of Philosophy
University of Western Ontario
askelto4@uwo.ca

ABSTRACT

Sidgwick famously claimed that an argument in favour of utilitarianism might be provided by demonstrating that a set of defensible philosophical intuitions undergird it. This paper focuses on those philosophical intuitions. It aims to show which specific intuitions Sidgwick endorsed, and to shed light on their mutual connections. It argues against many rival interpretations that Sidgwick maintained that six philosophical intuitions constitute the self-evident grounds for utilitarianism, and that those intuitions appear to be specifications of a negative principle of universalization (according to which differential treatments must be based on reasonable grounds alone). In addition, this paper attempts to show how the intuitions function in the overall argument for utilitarianism. The suggestion is that the intuitions are the main positive part of the argument for the view, which includes Sidgwick's rejection of common-sense morality and its philosophical counterpart, dogmatic intuitionism. The paper concludes by arguing that some of Sidgwick's intuitions fail to meet the conditions for self-evidence which Sidgwick himself established and applied to the rules of common-sense morality.

0. One aim of Henry Sidgwick's *The Methods of Ethics* is to provide an argument for utilitarianism, the view that an agent acts rightly insofar as she performs that action, out of the range of actions open to her, which maximizes aggregate happiness, hedonistically construed. He takes intuitions to be central to this aim. He maintains that 'the utilitarian method…could not…be made coherent and harmonious without…[a] fundamental intuition' (ME xvi-xvii), that 'the only moral intuitions which sound philosophy can accept as ultimately valid are those which at the same time provide the only possible philosophical basis of the Utilitarian creed' (PC 564), and that 'the Intuitional method rigorously applied yields as its final result the doctrine of pure Universalistic Hedonism, – which it is convenient to denote by the single word, Utilitarianism' (ME 406-407).[1] The

[1] For the abbreviations used herein, see the bibliography of primary sources below.

nature and number of intuitions and the role that they play in Sidgwick's argument is obscure. My purpose here is to clarify his position. In §§ 1 & 2, I defend an account of the nature and number of intuitions on which he relies. In § 3, I attempt to make sense of how the intuitions function in the argument for utilitarianism. In § 4, I briefly outline some worries about the intuitions.

1. Sidgwick subscribes to philosophical intuitionism, the view that there are 'one or more principles more absolutely and undeniably true and evident' (ME 102). These principles are self-evident: a proper understanding of them is sufficient for justifiably believing them (ME 229). The justification of these principles is therefore direct or arrived at by 'direct reflection' on the nature of the propositions in question (ME 383), though no intuition is infallible (ME 211; cf. ME 400). He calls this position intuitional in the 'wider sense' because with other intuitional positions it shares a commitment to 'self-evident principles relating to "what ought to be"' (ME 102n1). Intuitionism in the 'narrower sense' is dogmatic intuitionism. It is committed to the existence of self-evident propositions which are general rules 'implicit in the moral reasoning of ordinary men, who apprehend them adequately for most practical purposes' (ME 101). More specifically, it claims that 'we have the power of seeing clearly that certain kinds of actions are right and reasonable in themselves, apart from their consequences; – or rather with a merely partial consideration of consequences, from which other consequences admitted to be possibly good or bad are definitely excluded' (ME 200). The kinds of actions that are right are those required by the rules of justice, benevolence, and veracity, among others. A third species of intuitionism, perceptional intuitionism, holds that we intuit the morality of particular actions without reliance on rules or principles (ME 100).

Sidgwick rejects both dogmatic and perceptional intuitionism en route to his defense of philosophical intuitionism and utilitarianism. He does not devote much space to perceptional intuitionism but it is clear that he rejects it (ME 100-101, 214). The argument contra dogmatic intuitionism is more sustained and more central to his endorsement of philosophical intuitionism (ME 337-361). After an exhaustive survey of the various rules of common-sense morality with which the dogmatic intuitionist is concerned, he argues that we must reject the normative aspect of the view on the grounds that none of the rules, 'when fairly contemplated, even appears to have the characteristic of a scientific axiom' (ME 360). The problem is that the rules of common-sense morality are

unclear, or if clear, then disputed, or in conflict with each other, and therefore do not satisfy the four conditions of self-evidence, which require that for a proposition to be self-evident it must be 'clear and precise', 'ascertained by careful reflection', consistent with other propositions considered self-evident, and disagreement regarding its truth be absent or explained away (ME 338-342). At most, the rules of common-sense morality provide adequate guidance to typical people in typical circumstances. In the wake of his rejection of dogmatic intuition Sidgwick finds 'certain absolute practical principles, the truth of which, when they are explicitly stated, is manifest; but they are of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case; particular duties have still to be determined by some other method' (ME 379). These philosophical intuitions provide 'a rational basis for the Utilitarian system', the method by which we determine our particular duties (ME 387; see also ME 406-407).

Sidgwick relies on the following six philosophical intuitions.

1. 'It cannot be right for *A* to treat *B* in a manner in which it would be wrong for *B* to treat *A*, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment' (ME 380). Call this intuition U.

2. 'The mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment that [*sic*] to that of another' (ME 381). Call this intuition T.

3. 'The good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other; unless, that is, there are special grounds for believing that more good is likely to be realized in the one case than in the other' (ME 382). Call this intuition P.

4. 'As a rational being I am bound to aim at good generally, – so far as it is attainable by my efforts, – not merely at a particular part of it' (ME 382). Call this intuition B.

5. 'Happiness (when explained to mean a sum of pleasures)… [is] the sole ultimate end' (ME 402; see also LE 107, 128-130). Call this intuition H.

6. 'The greater *quantum* of pleasure is to be preferred to the less, and that *ex vi termini* the larger sum made up of less intense pleasures is the greater quantum of pleasure' (LE 110; italics in original). Call this intuition M.

I will now attempt to justify the contention that Sidgwick relies on six philosophical intuitions. He thinks U is self-evident: immediately preceding it he says that 'the self-evident principle strictly stated must take some such negative form as this' (ME 380). This proposition requires unpacking. It entails that one be consistent in one's moral judgements. If one claims that a certain act x is wrong, then one is rationally bound to claim that act y is wrong if x and y are identical in all their universal properties, i.e., features that may be stated as reasonable grounds for differentiating moral treatment or assessment. But what constitutes a 'reasonable ground' for variation in evaluative assessment? In discussing the intuition only unreasonable grounds appear to be discussed. This is not surprising: the axiom is 'negative', intending to 'throw a definite *onus probandi* on the man who applies to another a treatment of which he would complain if applied to himself' (ME 380). One ground that is explicitly ruled out as unreasonable is one that appeals to properties explicated purely in terms of particulars, i.e., non-generic terms.[2] The intuition, it seems, is intended to rule out as reasonable grounds such items as numerical differences, proper names and indexical terms, spatial location, essential reference to individuals, and so on. The intuition requires consistency in one's moral judgements with variations based on reasonable grounds alone, where reasonable grounds exclude non-generic terms.[3]

In his initial discussion of T Sidgwick does not say that the principle is self-evident. Instead, he implies it by stating that T is another 'principle' epistemologically analogous to U (ME 380-381). But only two pages later he

[2] For an excellent discussion of this issue, see Michael Smith, "Does the Evaluative Supervene on the Natural?," *Well-being and Morality: Essays in Honour of James Griffin*, eds., Roger Crisp and Brad Hooker (Oxford: University Press, 2000), 91-114, esp. 97-101, & J. L. Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin, 1977), 83-102.

[3] This intuition does not fill this notion out completely, however. This may leave Sidgwick open to the charge that this intuition is not clear and precise, though see below for more on this.

confirms that he holds that T is self-evident (ME 383). T and the other intuitions attempt to build on U. Each specifies further both 'unreasonable' and 'reasonable' grounds for varying one's moral judgements. In an early paper Sidgwick confirms this. 'The essence of Justice or Equity, in so far as it is absolutely obligatory, is that different individuals are not to be treated differently, except on grounds of universal application: which grounds, again, are given in the principle of Rational Benevolence' (UG 31). T expresses the idea that location in time is not directly or intrinsically relevant to the value of a state of affairs or experience.[4] T requires that agents remain rationally indifferent to when benefits and burdens occur. Sidgwick provides what look like several different versions of T, e.g., that 'I ought not to prefer a present lesser good to a future greater good' (ME 383) and that 'a smaller present good is not to be preferred to a greater future good' (ME 381), though these remain consistent with T.

P is described as a 'self-evident principle' (ME 382) and B is characterized as 'evident' and as a 'rational' intuition (ME 382) and as self-evident (ME 383). P and B give expression to some of the central features of utilitarianism. The general upshot of accepting them is that to whom a benefit or burden accrues is not directly significant to the morality or rationality of action. P is designed to nudge us towards this position by abstracting from one's own identity and adopting the point of view of no one in particular. From this viewpoint – the 'point of view…of the universe', as he calls it – we notice that each person has a good but that no one person's good is of more importance than another person's good. In taking up this point of view Sidgwick finds it self-evident that no one person's good satisfies what we might call a 'uniqueness condition', a condition the satisfaction of which would make it special and therefore more intrinsically important than another person's good. The exclusive role of P in the establishment of utilitarianism is that it opens up the possibility for a radically impartial theory of rational action, though it is important to point out that P presupposes that it is possible to compare the goods of individuals as against each other. Sidgwick explicitly states that the only legitimate ground for giving one person's good more attention is if that good happens to be greater than

[4] A factor, that is, independent of the quantity of goodness under consideration. For example, if x and y are of equal goodness in terms of their quantity, then other things being equal we ought rationally to have equal regard for them, despite the fact that the occurrence of x takes place at time t while y takes place at time t+ 1 year.

others in the class of all those being compared, and this implies comparability of the good and hence the possibility of aggregating goods both interpersonally and intrapersonally.

What is not implied by P is anything about how rational beings are required to act. From the fact that when viewed from the point of view of the universe it is possible to compare goods across individuals and to discover that no one's good is any more important than another's, it does not follow that we should be impartial as regards individual goods or that we should promote the good impartially construed. It is possible to grant that my good is of no more importance than another's, but hold that rationally speaking we have only to promote our own good on the whole. Similarly, it is possible to grant the claim about the possibility of comparability, of commensurability and of aggregation, but hold that rationally speaking we have only to promote our own good on the whole. This is, I think, something Sidgwick would accept, since for him the real debate between the egoist and the utilitarian turns on whether reasons are agent-relative rather than agent-neutral or vice versa (ME 420). This is what makes B key to the debate between rational egoism and utilitarianism, and it is clear that he regards is as such (ME 387-388, 500).[5] B represents the agent-neutrality that is at the very heart of utilitarian moral theories. It claims that the fact that something is good gives anyone and hence everyone a reason to desire or promote it. The mere fact that an act (or whatever) advances the good gives anyone a reason to do it.

Sidgwick thinks that a maxim of benevolence follows from P and B. As he puts it: 'from these two rational intuitions we may deduce, as a necessary inference, the maxim of Benevolence in an abstract form: viz. that each one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him' (ME 382). The precise manner in which this proposition follows from P and B is unclear. The claim that one ought to regard the good of another as much as one's own is misleading, for what happens when one does not have any regard for one's own good? It is better to construe the inference as stating that one is bound to maximize the good no matter whose it happens to be, since his view is that we do find something of value from the point of view of the universe. This makes it

[5]Rational egoism is the view that an agent is rational insofar as he seeks to maximize his own happiness, hedonistically construed (ME 95, 121).

consistent with B, which enjoins promotion of the general good rather than enjoining parity in treatment between oneself and others. The inference from P and B could be stated as follows: as a rational being I am bound to aim at good generally unless there is some sort of non-arbitrary reason not to do so. That is, I am bound to aim at good generally unless someone's or a group's good satisfies something like the uniqueness condition. From the point of view of the universe it appears self-evident that no one person's good is of any more importance than another person's, other things being equal. Therefore, from a denial of the fact that anything satisfies a uniqueness condition together with the claim that I have reason to aim at the good generally, it follows that I am morally bound to aim at the good, agent-neutrally construed.

Why does Sidgwick call this a necessary inference? The only real change between B and it is (at least in the way I have construed it) in the use of the phrase 'as a rational being I am bound' in B and the use of the phrase 'each one is morally bound' in the necessary deduction. This is a necessary inference because for him "rationally bound" is synonymous with "morally bound" (ME 375, 34-35). Shortly after completing his account of U, T, P, B and the deduction, Sidgwick maintains that he has arrived, 'in my search for really clear and certain ethical intuitions, at the fundamental principle of Utilitarianism' (ME 387). But as he notes this is not quite accurate, since 'to make this transition logically complete, we require to interpret "Universal Good" as "Universal Happiness"' (ME 388).

It is not obvious that H is self-evident. At best U, T, P, and B together establish some sort of maximizing consequentialism.[6] However, Sidgwick's remarks indicate that he wants to establish utilitarianism using the intuitional method, not just maximizing consequentialism (ME xvi-xx, 387, 388, 406-407, UG 31-33, PC 564). If we are to take this claim seriously, we need to consider whether or not he actually thinks there is an intuition pertaining to the ultimate good. Without such an intuition it seems that we cannot make sense of his claim to have established utilitarianism by the intuitive method.

In defending his claim that happiness is the only thing good in itself, Sidgwick asks that 'the reader…use the same twofold procedure that I before requested him to employ in considering the absolute and independent validity of common moral precepts' (ME 400). The twofold process involves both an appeal

[6] J. B. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy* (Oxford: Clarendon Press, 1977), 304.

to 'intuitive judgement [of the reflective intellect] after due consideration of the question when placed fairly before it' and 'a comprehensive comparison of the ordinary judgements of mankind' (ME 400; see also LE 127).[7] As regards the first procedure Sidgwick states that upon 'sober' reflection in, as Butler says, 'a cool hour', he arrives at the following intuition: 'we can only justify to ourselves the importance that we attach to any of these objects ['Virtue, Truth, Beauty, Freedom' (ME 400)] by considering its conduciveness, in one way or another, to the happiness of sentient beings' (ME 401). Elsewhere Sidgwick is more explicit: 'My own answer to the question…Why is the ultimate good and criterion held to be pleasure? is, that nothing but pleasure appears to the reflective mind to be good in itself, without reference to an ulterior end; and in particular, reflection on the notion of the most esteemed qualities of character and conduct shows that they contain an implicit reference to some other and further good' (LE 107). Furthermore, he says that he appeals to intuition to 'justify my own view that it is Pleasure alone, desirable Feeling, that is ultimately and intrinsically good' (LE 126). The intuition appears to be that happiness which consists in pleasure defined as 'a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable or – in cases of comparison – preferable' (ME 127; see also ME 131, LE 130) is the sole ultimate good.[8] That he thinks he has obtained an intuition with respect to the ultimate good explains (a) why he thinks that the intuitional method when rigorously applied

[7]He is not here trying to justify his account of ultimate good by reference to common sense itself. He says that his aim is to '*bring* Common Sense to this admission [namely]…that Happiness is the only thing ultimately and intrinsically Good or Desirable' (ME 421n1; italics added). This may not always have been the case. In an early discussion he argues that to establish his view of ultimate good he appeals to 'the immediate intuition of reflective persons; and…to the results of a comprehensive comparison of the ordinary judgements of mankind' (UG 35). Here he contends that the argument from ordinary judgements 'comes in rather by way of confirmation of the first' (UG 35), suggesting that he thinks the ordinary-judgements argument confirms the intuitive one. In the final edition of ME, however, he drops the claim about confirmation, which suggests that he changed his mind on this point (see also LE 128). Whatever the case may be, in both cases it looks like he is employing the intuitional method (at least in part) to arrive at an account of ultimate good. For an account of Sidgwick's attitude toward the epistemological status of common-sense morality, see Anthony Skelton, "Schultz's Sidgwick," *Utilitas* 19 (2007), 91-103.

[8] It is not obvious what Sidgwick means by pleasure. The passage quoted in the text is just one of the accounts of pleasure he provides. For others, see ME 94; 93 & 120-121; & 402. He appears to favour the account quoted in the text; see LE 130, ME 398.

leads to utilitarianism and (b) his frequent appeals to intuitive reflection in his discussion of the nature of the ultimate good.

Of course Sidgwick nowhere declares explicitly that H it is self-evident. He argues only that he relies on the intuitional method to arrive at H. But his only account of what the intuitional method comprises suggests that he holds that H is self-evident. Recall his account of an intuition: 'by calling any affirmation as to the rightness or wrongness of actions "intuitive," I do not mean to prejudge the question as to its ultimate validity, when philosophically considered: I only mean that its truth is apparently known immediately, and not as the result of reasoning' (ME 211; see also PC 564). If this is a basic feature of the intuitional method, then we may conclude that Sidgwick arrives at his account of the ultimate good in the same way that he arrives at his other intuitions, by direct reflection on the proposition in question. My suggestion is confirmed by his only other explicit discussion of the relationship between hedonism and intuitionism (ME 98). He claims that hedonism is authoritative just in case happiness, hedonistically construed, is the ultimate reason for action. This claim is not known by induction from experience in the way Mill might have thought. Rather, if the claim that pleasure is the only reasonable ultimate end of human action 'is legitimately affirmed in respect either of private or of general happiness, it must either be immediately known to be true, – and therefore, we may say, a moral intuition – or be inferred ultimately from premises which include at least one such moral intuition; hence either species of Hedonism, regarded from the point of view primarily taken in this treatise, might be legitimately said to be in a certain sense "intuitional"' (ME 98).[9] Since he does not infer his own account of ultimate value from premises it must be the case that he thinks it is known by intuition, hence he thinks it is self-evident that the good is happiness, hedonistically construed, for a moral proposition is a moral intuition only if it is self-evident.

Sidgwick thinks he arrives at a maximizing version of utilitarianism (ME 411). It is not made explicit how he gets maximization out of his intuitions. His thought might be that doing less than the maximum would result in aiming at only part of the good. By doing less than the maximum one would be aiming merely at a particular part of the good rather than at good generally. However, it looks like Sidgwick gets maximization in another way. He does not claim explicitly that we ought to maximize the good. Instead, he seems to think that

[9] He is referring here to the wider sense of intuitional; see ME 98n2.

if 'it be granted that pleasure as the end is made up of elements capable of quantitative comparison', then it is 'self-evident' that M (LE 110). Together B, H and M require that we aim at maximal happiness or pleasure agent-neutrally construed.

2. I have argued that Sidgwick relies on six philosophical intuitions in his argument for utilitarianism. In this section I argue against rival interpretations.

In *The Theory of Good an Evil*, Hastings Rashdall suggests that Sidgwick relies on three philosophical intuitions.[10] He maintains that Sidgwick holds it self-evident that 'I ought to promote my own good on the whole (where no one else's good is affected), that I ought to regard a larger good for society in general as of more intrinsic value than a smaller good, and that one man's good is (other things being equal) of as much intrinsic value as any other man's.'[11] He calls these prudence, rational benevolence and equity. He misses U, T, M and H. He might be forgiven for missing H, but not for missing the others, which are clearly labeled self-evident.[12] Sidgwick does not in ME state that Rashdall's 'prudence' is self-evident. At best such a requirement falls out of the requirement to advance the aggregate good in a case where one finds oneself marooned on an uninhabited desert island. Rashdall must be confusing 'prudence' with T.[13] His rational benevolence and equity resemble B and P and the necessary inference discussed above. Nevertheless, he misses the key element of B, namely, that we ought to aim at good generally rather than at merely a particular part of the good: he gives the intuition an axiological, rather than

[10] Hastings Rashdall, *The Theory of Good and Evil*, Vol. I (Oxford: University Press, 1907), 90-91, 147, & 184-185.
[11] Rashdall, 90-91.
[12] Rashdall is aware that a claim like H relies on intuition for justification. He believes that Sidgwick relies on intuition to justify something like H, though he does not seem to think that Sidgwick thinks that H is self-evident and therefore on the same level as the three other self-evident intuitions that Rashdall lists. See *Ethics* (London: T. C. & E. C. Jack, 1913), 22, and see also *Is Conscience an Emotion?* (Boston: Houghton Mifflin, 1914), 43.
[13] Or, he may be misled by Sidgwick's sometimes sloppy account of his intuitions; see ME 391-392, FC 483.

deontic gloss.[14] The deduction is not itself self-evident; it is deduced from self-evident propositions.

J. B. Schneewind agrees that Sidgwick endorses U, T, P and B.[15] He misses M. He argues that Sidgwick gets maximization from 'the definitions of rightness and goodness.'[16] This is untrue. First, Sidgwick claims that 'right' is not definable (ME 32, 32-33, FC 480). Second, he holds that 'good' is definable, but in his definition he does not mention the idea of maximization (ME 112). Third, he is keen to ensure that definitions of key moral terms (e.g., right and ought) remain neutral with respect to substantive moral questions (FC 480-483, ME 109). Therefore, he is unlikely to be warm to the idea of getting maximization from definitions of central moral and axiological notions.

Schneewind's point might be understood in another way. When he refers to 'definitions' he might be referring to the way in which B connects the right and the good.[17] He claims that what demonstrates that maximal goodness is what makes acts right is 'the negative result of the examination of common-sense morality, that none of the purely factual properties of acts can serve as an ultimate right-making characteristic. It cannot, therefore, be the case that some factual properties of acts make them right…it must rather be the case that bringing about the most good is what makes right acts right.'[18] This is difficult to swallow. One might grant the results of the negative argument against common-sense morality and that the good is the ultimate-right making characteristic as per Schneewind's account of the intuitions, but deny that it is maximal goodness that is the ultimate right-making characteristic. It is still the case that from B one needs an argument or something analogous to get one to the claim that we ought to maximize the good, rather than simply promote it to some degree. Indeed, Sidgwick seems required to run an argument analogous to

---

[14] Rashdall discusses Sidgwick's intuitions again in *Ethics*, where he gives rational benevolence a deontological gloss; see *Ethics*, 62. In the same place, however, he construes prudence as the claim that 'I ought to promote my own greater good rather than my own lesser good' and rational benevolence as the claim that 'I ought to promote the greatest good on the whole' (62). As noted, Sidgwick does not defend prudence in ME, and these two intuitions sometimes conflict.

[15] Schneewind, 296. For a similar account of Sidgwick's intuitions, see Robert Shaver, *Rational Egoism* (Cambridge: University Press, 1999), 61-62, 74.

[16] Schneewind, 307.

[17] I owe this suggestion to Robert Shaver.

[18] Schneewind, 308.

the one he runs against common-sense moral rules against rivals of the maximizing conception of rationality.

On the face of it, Schneewind does not think that Sidgwick endorses an intuition pertaining to the ultimate good. Officially, his position appears to be that Sidgwick embraces *only* U, T, P and B.[19] This is not satisfactory. This would leave the view of the ultimate good undefended in ME, and it is not consistent with his defense of his account of ultimate good elsewhere (e.g., LE 107, 126ff.). Sidgwick also maintains that the justification of a claim like H is either inferential (i.e., inferred from a set of propositions which include at least one intuition) or intuitive (ME 98). It appears not to be inferred from any set of propositions which include at least one intuition; therefore, it must be justified by reference to intuition.

Schneewind appears to suggest that he believes this. He claims that although in Book III, chapter XIV of ME Sidgwick maintains that there is no self-evident principle 'enabling us to connect ultimate good out of all relation to consciousness with human action'[20] and he is 'not appealing to an additional intuition to exclude the intrinsic goodness of things or states of affairs out of relation to all consciousness, but is asserting only that he finds no self-evident practical principle asserting their goodness', he does defend the 'utilitarian principle'.[21] By this he means that Sidgwick has 'not just one axiom – that pleasure is intrinsically good – but as many self-evident propositions as there are experiences of pleasure.'[22] In his case, Schneewind's position is that there are the four intuitions that he explicitly notes, plus an intuition pertaining to the good, namely, that pleasure is intrinsically good, plus as many as there are experiences of pleasure.[23] This is problematic. First, this conflicts with Schneewind's interpretive requirement that 'it seems sensible to try to find the smallest number of axioms with which the work to be done by first principles can be done.'[24] Second, when Sidgwick discusses the intuitive argument for his account of the ultimate good he refers to pleasure as he defines it as being the only thing

[19] Schneewind, 290.
[20] Schneewind, 325.
[21] Schneewind, 326.
[22] Schneewind, 320.
[23] Schneewind provides no argument for the general claim about the value of pleasure, which leads me to believe that he does not think that there is such an intuition. His focus is entirely on showing that claims about the value of particular pleasures are self-evident.
[24] Schneewind, 290.

that is ultimately good or to the claim that all and only the happiness, hedonistically construed, of sentient beings possesses ultimate goodness (e.g., LE 107, 126ff., ME 402, 398). He does not say that certain particular feelings are themselves self-evidently desirable. He seems to think that the general claims about pleasure or happiness are self-evident.[25] Schneewind is misled here by Sidgwick's view of pleasure. The latter defines pleasure as 'a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable or – in cases of comparison – preferable' (ME 127). It seems that Schneewind believes that the variety of apprehension mentioned here is intuitive in nature. The position is that each feeling of the sort that Sidgwick picks out is intuitively known by the one experiencing it to be desirable or intrinsically valuable. He appears at times to use apprehension in this way (ME 383). However, it is not obvious from anything he says that he intends to use it in this way in his definition of pleasure. The fact that one's apprehension of the desirability of certain feelings is not likely to be arrived on the basis of understanding alone and the fact that Sidgwick believes that non-human animals can experience pleasure indicates that he does not intend to use the term this way.[26]

J.M.E. McTaggart argues that Sidgwick produces five intuitions.[27] Unlike other commentators, McTaggart is aware that Sidgwick has an intuition resembling H, though he provides no argument for this.[28] My argument above vindicates his assertion. However, he holds that there is another axiological intuition, that 'nothing…is good as an end except some state of a conscious being; and nothing is good as a means except as tending to bring about some state of a conscious being.'[29] There is some evidence that this is Sidgwick's view. At the conclusion of his discussion of the notion 'good' he says that 'we can find nothing that, on

---

[25] Robert Shaver has suggested to me that the claim about particular pleasures may simply be an application of the general claim that pleasure is intrinsically valuable. In this case, however, it would be mistaken to think that claims about particular pleasures are axioms rather than derivations from an axiom and this cannot be Schneewind's view because he contends that the particular episodes of pleasure meet the tests that Sidgwick applies to self-evident intuitions (Schneewind, 319).

[26] For the claim about animals, see ME 414.

[27] J. Ellis McTaggart, "The Ethics of Henry Sidgwick," *Quarterly Review* 205 (1906), 398-419.

[28] For the same, see William Frankena, "Sidgwick, Henry," *An Encyclopedia of Morals*, ed., Vergilius Ferm (New York: Philosophical Library, 1956), 539-544, 542.

[29] McTaggart, 407.

reflection, appears to possess this quality of goodness out of relation to human existence, or at least to some consciousness or feeling' (ME 113; see also LE 124). Sidgwick often uses the language of reflection in his discussion of the intuitions above (see, e.g., ME 383). This seems to indicate that we should interpret him as holding this intuition. But in the case of the above intuitions in general and in the case of H in particular he says that he relies on intuition or that they are self-evident; he does not say this with respect to the claim that McTaggart refers to. He seems instead to treat this claim as a lemma in his argument for the proposition that happiness (hedonistically construed) is the sole ultimate good (ME 398). Moreover, since he raises objections to it, it is best to see him as holding that it does not qualify as an intuition.

McTaggart also maintains that Sidgwick thinks that it is self-evident that 'we ought to prefer the good to the bad.'[30] He lists no evidence that Sidgwick thinks this, and it seems more likely that Sidgwick thinks that it is part of the definition of good that we ought to seek it, and that it is part of the definition of bad that we ought not to seek it. Indeed, he defines 'ultimate good on the whole' as 'what as a rational being I should desire and seek to realize, assuming myself to have an equal concern for *all* existence' (ME 112; italics in original). McTaggart misses some of the other intuitions (e.g., P and M). Most surprising is the fact that he misses U. Sidgwick holds that U is self-evident. Immediately preceding U the following words appear: 'the self-evident principle strictly stated must take some such negative form as this' (ME 380). It may be that McTaggart is misled by the fact that Sidgwick is not entirely explicit about the status of this requirement. He sometimes treats U as a logical requirement built into the meaning of moral terms, and perhaps McTaggart's belief is that this is Sidgwick's considered view.

In one of his main discussions of meta-ethics Sidgwick claims that terms like 'ought' and 'right' and their cognates are 'too elementary to admit of any formal definition' (ME 32; see also FC 480-483). The only method by which to clarify the fundamental notion is 'by determining as precisely as possible its relation to other notions with which it is connected in ordinary thought' (ME 33). One 'notion' with which these terms are connected (and with which they are liable to be 'confounded') is the following. 'When a moral judgement relates primarily to some particular action we commonly regard it as applicable to any other action belonging to a certain definable class: so that the moral truth

[30] McTaggart, 408.

apprehended is implicitly conceived to be intrinsically universal, though particular in our first apprehension of it' (ME 34).[31] Although he maintains that this notion is intimately connected with 'ought' judgements he does not claim that the notion or requirement is built into the meaning of the term and its cognates. Indeed, he argues that these latter terms are not definable. This suggests that he is not a proponent of the logical thesis that universality is part of the meaning of moral judgements. This is further confirmed by his suggestion, when discussing the principle elsewhere, that the requirement of universality is 'implied in the common notion of "fairness" or "equity"' (ME 380), a substantive normative principle.[32] In addition, he treats this principle in the same way that he treats the other self-evident intuitions, namely, as requirements of rationality (ME 386-387).

There is some evidence, however, that he holds that it is a logical requirement contained in the meaning of moral notions. In his discussion of dogmatic intuitionism he claims that the following is obtained by merely 'reflecting on the general notion of rightness' (ME 208). 'We cannot judge an action to be right for *A* and wrong for *B*, unless we can find in the natures or circumstances of the two some difference which we can regard as a reasonable ground for difference in their duties' (ME 209). His remark that he finds this principle by reflecting on the notion of rightness suggests that he believes the requirement to be one of logic. But this is a little too quick. In the discussion mentioned in the last paragraph he says that the requirement of universality is 'connected in ordinary thought' with terms like 'right' and 'ought' despite not being part of the definition of these terms. In his later discussion he refers to the notion of rightness as 'commonly conceived'. This suggests that, although the requirement is found in the notion of rightness 'as commonly conceived', it is not strictly speaking part of the meaning of the term 'right' or 'ought'. This is, I think, the best way to reconcile his later comments with those discussed above. Finally, Sidgwick explicitly states that he wants to arrive at 'self-evident moral principles of real significance' (ME 379), not merely tautologies or 'sham-axioms' (ME 374). This gives us a strong reason to think that he does not intend the principle as a logical thesis, but as a self-evident principle.

---

[31] This resembles his final articulation of U; he explicitly connects the two at ME 208n2.

[32] By 'implied in' he does not mean built into the meanings of the term; see ME 386.

A.R. Lacey argues that Sidgwick espouses seven intuitions.[33] His account of the intuitions is close to mine, though he, too, misses M and H. He mistakenly lists the necessary inference as an intuition.[34] He lists U, but he thinks that there are further intuitions with respect to justice in ME. According to Lacey, Sidgwick holds that the following two claims are self-evident. 'If a kind of conduct that is right (or wrong) for me is not right (or wrong) for some one else, it must be on the ground of some difference between the two cases, other than the fact that I and he are different persons' and that we ought to exhibit 'Impartiality in the application of general rules'.[35] Sidgwick discusses both of these requirements. As regards the first, he says that it is 'widely recognized', but after raising objections to it and some other similar accounts he says that the self-evident principle 'strictly stated' is U. The others are either imprecise or applications and he accepts U in part because it is precise (ME 380). Of the second requirement, Sidgwick does say that there 'appeared to be no other element which could be intuitively known with perfect clearness and certainty' (ME 380). The key word here is 'appeared'. It may be the case that it appeared to be that there was no other element which could be known intuitively, but Sidgwick's view seems to be that the appearance is illusory, since he claims that there are no self-evident propositions to be found in common-sense morality (ME 360). At best, the requirement is another formulation of U.

Some further matters need to be dealt with. T is often regarded as the basis for rational egoism.[36] However, as T stands here it is consistent with both rational egoism and utilitarianism.[37] The intuition does not tell one whether or

---

[33]A. R. Lacey, "Sidgwick's Ethical Maxims," *Philosophy* 34 (1959), 217-228. For a similar account of Sidgwick's intuitions with some of the same errors, see C. D. Broad, *Five Types of Ethical Theory* (London: Kegan Paul, 1930), 223-227.

[34] Lacey, 219. It may be that Rashdall, McTaggart, Lacey and others are mislead on this score by previous editions of ME, where Sidgwick lists the 'necessary inference' found in ME as an intuition (see ME2 355, ME3 381-382, ME4 382). This mistake is also found in F. H. Hayward, *The Ethical Philosophy of Sidgwick* (London: Swan Sonnenschein, 1901), 110. Hayward notes that Sidgwick endorses T and U.

[35] Lacey, 218.

[36]Schneewind, 362, and Bernard Williams, "The Point of View of the Universe: Sidgwick and the Ambitions of Ethics," *Making Sense of Humanity* (Cambridge: University Press, 1995), 153-171, 160-161.

[37] For this point, see Georg von Gizycki's review of ME4, *International Journal of Ethics* 1 (1890), 120-121. See also Shaver, 75, Schneewind, 361, and Hastings Rashdall, "Professor Sidgwick's Utilitarianism," *Mind* 10 (1885), 200-226, esp. 202, and ME 414.

not one should give greater regard to one's own good on the whole than the good on the whole of others.[38]

But Sidgwick is not always careful. He claims at one point that prudence, 'so far as…[it is] self-evident, may be stated as…[a precept] to seek…one's own good on the whole, repressing all seductive impulses prompting to undue preference of particular goods' (ME 391-392). There are three reasons for thinking that his considered view is that only T is self-evident. First, he could not have intended to argue for something that would lead to egoism in his discussion of T, for this directly contradicts his claim that the intuitions he discusses in ME (U, T, P, B and H) provide a 'rational basis' for utilitarianism, not both utilitarianism and egoism (ME 387). If the intuition is supposed to refer not only to T but also to the essential features of rational egoism, these claims about establishing utilitarianism are baffling at best. Second, by his own account, he did not attempt to establish the truth of egoism in the first three editions of ME (FC 484). However, starting in the second and third editions T and some other intuitions that pertain to utilitarianism *are* present (ME2 354, ME3 380-381). If the intuition did prove rational egoism, he could not say that he provided no argument for it in the second and third editions. Sidgwick does say in ME3 and ME4 that T is the 'principle on which…Rational Egoism is based' (ME3 388, ME4 386-387). But he noticed that this conflicted with his claim not to be providing a basis for the view. Hence, in subsequent editions he stated very clearly that T is merely 'implied in' rational egoism (ME 386). It is not there providing a 'rational basis' for egoism in the way that B provides (or appears to provide) a rational basis for utilitarianism (ME 387).[39] Third, when he does turn to a discussion of what the basis of rational egoism might be, he does not refer to T. Instead, he contends that the 'rationality of Egoism is based [on]…the assumption…that the distinction between any one individual and any other is real and fundamental, and that consequently "I" am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals' (FC 484; see also ME 498). He declares that this proposition is the 'self-evident' intuition 'upon which the rationality of Egoism is based' (FC 484).

---

[38] Indeed, even the practical manifestation of the principle is agnostic as to whether individual or aggregate good is to be promoted.

[39] For this point, see Shaver, 76.

It is important to note here that Sidgwick may be construed as producing a seventh intuition in his discussion of the basis of rational egoism. I dissent from this construal. First, Sidgwick does not declare that the above proposition is self-evident in ME (ME 498). Second, there is good reason for this and hence good reason for thinking that this is not a seventh intuition. This intuition appears to fail the clarity and distinctness test. It runs the idea of separateness of individuals together with the claim about its role in our thoughts about what we have most reason to do. Precision requires separating various ideas from each other and this putative intuition does not achieve this.[40] It might be that in ME Sidgwick means to connect the claim about separateness with reasons for action. This seems problematic. The passage states that the second claim follows from the first but this is not an obvious or necessary truth, for while there may be cases where this is true, there are well known counter-examples. Mother Teresa might well have noted that she is a separate individual with a set of projects and commitments that drove only her as any agent, but that she was not concerned with her own good in a way that was fundamentally more important and different than her concern for others. It is also not obvious just how fundamental the unconcern for others is that follows from the distinctness. I might be concerned with myself in a way more fundamental than the way I am concerned with you but still hold that I have at times a duty to help others, for example, where the cost to me is negligible and the benefit to you is great. I might think that for the most part I am concerned for myself but not entirely. If this proposition is to pass the test and get us to rational egoism it has to construe 'fundamental' in the strongest possible sense. But this is unclear from the way the proposition is stated. In light of these problems and the fact that he does not declare that this claim is self-evident in ME, it seems best to think that Sidgwick's considered view is that there is not a seventh philosophical intuition.

3. My aim to this point has been to outline the philosophical intuitions on which Sidgwick relies in his argument for utilitarianism. But how does he demonstrate the truth of utilitarianism by reliance on the intuitions? Nowhere is any kind of deduction or argument from the intuitions as premises to utilitarianism as a conclusion provided. In this section I provide a schematic statement of how the

[40] A 'distinct notion of any object…[is] one that is not liable to be confounded with that of any different object' (LK 449).

intuitions figure into the argument for utilitarianism. The best way to see the role that the intuitions play in his argument is to situate them in the general structure of ME. The intuitive argument for utilitarianism forms one part of the argument for the view, which includes a negative argument contra common-sense morality and its philosophical counterpart, dogmatic intuitionism, the main features of which are found in Book III, chapter XI, the appeal to philosophical intuitions, which takes place primarily in Book III, chapters XIII & XIV, and a Millian-style proof, which is supplied in Book IV, chapters II & III.[41]

Had Sidgwick attempted an explicit argument, it might have looked as follows:

P1. As a rational being I am bound by the basic requirements of reason.

P2. The basic, ultimate requirements of reason direct one to do either what is based on what is right without reliance on all of the consequences that flow from what one is doing or on what is good without restriction (ME 2-3, 391, UG 27-28, OHE 6-7).

P3. It is not the case that the basic requirements of reason direct one to do what is right without reliance on all of the consequences that flow from what one is doing (ME 337-361). Instead, the morality of common sense is at best 'perfectly adequate to give practical guidance to common people in common circumstances' (ME 361).[42]

C1. Therefore, as a rational being I am bound to regard what is good without restriction (ME 391).

P4. Variation of treatment of individuals must be based on *reasonable* grounds alone, where this is considered to exclude non-generic grounds (ME 380). (This is U.)

P5. It is not reasonable to regard the time at which the good occurs as directly (or intrinsically) relevant to its value (ME 381). (This is T.)

P6. It is not reasonable to regard to whom the good accrues as directly (or intrinsically) relevant to the rationality of an action (ME 382; see also UG 31). Instead, one is required to advance the good, agent-neutrally construed. (This is a combination of P and B.)

---

[41] For more on the nature of the Millian proof, see Henry Sidgwick, "The Establishment of Ethical First Principles," *Mind* 4 (1879), 106-111.

[42] This is the conclusion of the negative argument against common-sense morality and dogmatic intuitionism.

P7. 'Happiness (when explained to mean a sum of pleasures)…[is] the sole ultimate end' (ME 402; see also LE 107), where pleasure is defined as 'a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable or – in cases of comparison – preferable' (ME 127; see also ME 131, LE 130). (This is H.)
P8. 'It is self-evident that the greater *quantum* of pleasure is to be preferred to the less, and that *ex vi termini* the larger sum made up of less intense pleasures is the greater quantum of pleasure' (LE 110; italics in original). (This is M.)
C2. Therefore, I, as a rational being, am morally bound to advance to a maximum degree happiness, agent-neutrally and temporally neutrally construed.
P9. If a method of ethics embodies or gives the best or most reasonable expression of these ultimate requirements of reason or intuitions, then it is true.
P10. Utilitarianism is the only method of ethics (that we know of) that embodies or is the best expression of these intuitions.
C3. Therefore, utilitarianism is the only method of ethics or rational procedure by which I determine what I ought to do.
C4. Therefore, as a rational being I am bound by the dictates of utilitarianism.

This seems a reasonable summary of the main argument for utilitarianism in ME, and of how the intuitions function in the argument. The intuitions provide epistemic justification for utilitarianism and emerge in the context of an argument against the claim that there are self-evident intuitions within common-sense morality, and this argument is supplemented by the Millian-style proof of Book IV, chapters I & II.

4. As I mentioned above, Sidgwick rejects the claim that the main rules of common-sense morality (e.g., justice, good faith, veracity and purity) are properly characterized as self-evident. Instead, his view is that 'such rules…are only valid so far as their observance is conducive to the general happiness' (ME 8). His main criticism is that the rules of common-sense morality fail his tests for self-evidence (discussed in § I) (ME 338-342). Broadly speaking, he argues that 'so long as they are left in the state of somewhat vague generalities…we are disposed to yield them unquestioning assent…But as soon as we attempt to give them the definiteness which science requires, we find that we cannot do this without abandoning the universality of acceptance' (ME 342). Sidgwick is very scrupulous when examining the putative intuitions of rivals; however, he is

much less than rigorous when it comes to demonstrating that his own intuitions satisfy the conditions for self-evidence. In this section, I briefly outline how some of his intuitions appear to fail the tests.

If we apply the clarity and precision and disagreement tests to Sidgwick's intuitions we do indeed find difficulties. In his discussion of U he does not define what we he means by a 'reasonable' ground for difference of treatment. The notion admits of several different interpretations, and although I have tried to clarify it, it is not clear or precise from examining U alone that all rational inquirers will agree on how to understand the notion of 'reasonable' or what is implied by it. If one examines B, one finds Sidgwick arguing that we ought to aim at good generally, not merely at a particular part of it. One might agree to this claim when it is put in this way. However, disagreement might emerge because rational inquirers have different views about how we ought to aim at the good. Some rational inquirers may well agree to B but only when it is accepted that the only appropriate way to aim at the good is directly rather than indirectly; other rational inquirers may agree to B but only when it is assumed that it is permissible to aim at the good directly and/or indirectly depending on what various empirical calculations dictate. Or one may agree to B when the good is left unspecified but reject it when the good is understood to consist in happiness or pleasure or some other good. Similar problems can be pointed out for P and various renditions of T where the notion of good is also left unspecified. T refers to the notion of consciousness. One might agree to T if consciousness is meant to include only higher-order consciousness, such as virtuous intending, intellectual activities, and the contemplation of beauty, but not if it is meant to include in addition all pleasure, feelings or emotions that do not require a kind of higher-order awareness or consciousness.

By far the most controversial intuition is H. It is not always manifest what Sidgwick believes is self-evident. Is it self-evident that pleasure is the sole ultimate good or is it self-evident that happiness hedonistically construed is the sole ultimate good? The difference between these two is that in the first case it is pleasure that is intrinsically valuable and in the second case it is happiness that is intrinsically valuable and then argued to consist in pleasure. It seems that it is the second claim, but Sidgwick does not properly distinguish between the two. One might agree to the second claim as it is presented in H, but disagree when pleasure is defined in the way that Sidgwick suggests, as 'a feeling which, when experienced by intelligent beings, is at least implicitly apprehended as desirable

or – in cases of comparison – preferable' (ME 127; see also ME 131, LE 130). Or, one might even agree to the account of pleasure just given but only because one interprets 'intelligent beings' in a certain way. What is meant by an 'intelligent being'? Is this notion meant to include more than fully developed adult humans? If not, then certain individuals may agree with Sidgwick's claim. But if so, then others may disagree. At times, he substitutes 'sentient' for 'intelligent' in his definition of pleasure (ME 131, 398). This suggests that he means to include more than simply fully developed adult human beings, and this may lead some to agree to Sidgwick's claim but it may lead to some disagreeing, especially those who are loath to grant non-human animals moral standing.

Sidgwick, of course, notes that there is deep disagreement about some of his intuitions. He is in fact all too willing to note that the rational egoist rejects P and B and that he cannot convince the egoist of utilitarianism using the Millian-style of proof (ME 420). He ends the work with the dualism of practical reason: both rational egoism and utilitarianism present themselves as equally reasonable though conflicting requirements of reason. This conclusion raises a worry about how Sidgwick understands the relationship between disagreement and his philosophical intuitions. He seems to suggest that where there is disagreement and where we 'have no more reason to suspect error in the other mind than in my own', then 'reflective comparison between the two judgements necessarily reduces me temporarily to a state of neutrality' (ME 342). If this is the case, then why is he not reduced to a state of neutrality with respect to the intuitions that play a role in the justification of utilitarianism? This seems the more reasonable position to advocate than a dualism of practical reason, the generation of which relies on maintaining the truth of utilitarianism and the intuitions that undergird it. Sidgwick is therefore unclear on just what to do in light of disagreement, and to the extent that he is unclear his argument against dogmatic intuitionism is weakened.[43]

Sidgwick does not deal well with disagreement in other cases. For example, he is aware that many reject his theory of value (ME 401, LE 126). However, in

[43] Sidgwick is also unclear as to what the clarity and precision test demands. He appears to fault common-sense morality and dogmatic intuitionism for not producing clear and precise practical directives. However, he notes that his own intuitions fail to tell us what to do in particular cases and that they do not give us complete practical guidance (ME 379, 380). He does not claim that they are impugned as a result. This seems unfair to the proponents of common-sense morality and dogmatic intuitionism.

addressing T. H. Green's criticisms of his view, for instance, his tactic is to rearticulate his arguments for H and to raise several objections to Green's own view. Is this sufficient to show that he has more reason to suspect error in Green's mind than in his own? If this is what Sidgwick has in mind, then it is something that he needs to explain better. In his defense of the view of the good in ME he addresses worries that might be raised by adherents of common-sense morality (ME 402ff.), and he employs arguments to show how certain ideal goods (truth, freedom, virtue, and so on) might be understood from a happiness theorist's point of view. But all this shows is that the happiness theorist may be able to make some sense of these rival values; it does not demonstrate that the dissenters are wrong. It is not clear how this might explain away the dissent or show that Sidgwick has more reason to suspect error in the mind of his opponent than his own.

Adherents of Sidgwick's intuitive argument for utilitarianism will need to both clarify his intuitions and respond to critics of them if it is to be acceptable. It will not do to simply state without explanation, as Rashdall does in his endorsement of some of Sidgwick's intuitions, that they 'possess the clearness and definiteness and freedom from self-contradiction which other alleged intuitions so conspicuously lack.'[44]

5. Sidgwick's argument for utilitarianism involves appeal to a number of philosophical intuitions. The nature and number of such intuitions is a matter of scholarly dispute. I have argued that he appeals to six philosophical intuitions in attempting to justify utilitarianism. This appeal is part of his general argument for utilitarianism which includes both a negative argument against common-sense morality and its philosophical counterpart, dogmatic intuitionism, and a Millian-style proof which attempts to convince critics of utilitarianism by reliance on views that they already accept. His argument will not be acceptable until the philosophical intuitions receive further clarification and defense. In particular, Sidgwick and those inclined to defend his argument for utilitarianism must demonstrate that the intuitions themselves meet the requirements that he suggests all self-evident propositions must meet if they are

[44] Rashdall, *The Theory of Good and Evil*, I, 90.

to function as premises 'that lead us cogently to trustworthy conclusions' (ME 338).[45]

Primary Sources

FC "Some Fundamental Ethical Controversies," *Mind* 14 (1889), 473-487.
LE *Lectures on the Ethics of T. H. Green, Mr. Herbert Spencer, and J. Martineau*. Edited by E. E. Constance Jones. London: Macmillan, 1902.
LK *Lectures on the Philosophy of Kant and Other Philosophical Lectures and Essays*. Edited by James Ward. London: Macmillan, 1905.
ME *The Methods of Ethics*, seventh edition. London: Macmillan, 1907. References to the second, third and fourth editions (London: Macmillan, 1877, 1884, 1890) take the form "ME2", "ME3" or "ME4".
OHE *Outlines of The History of Ethics for English Readers*, sixth edition. Edited by Alban Widgery. London: Macmillan, 1931.
PC "Professor Calderwood on Intuitionism in Morals," *Mind* 1 (1876), 563-566.
UG "Hedonism and Ultimate Good," *Mind* 2 (1877), pp. 27-38.

Secondary Sources

Broad, C. D. *Five Types of Ethical Theory* (London: Kegan Paul, 1930).
Frankena, William. "Sidgwick, Henry," *An Encyclopedia of Morals*, ed., Vergilius Ferm (New York: Philosophical Library, 1956).
Hayward, F. H. *The Ethical Philosophy of Sidgwick* (London: Swan Sonnenschein, 1901).
Lacey, A. R. "Sidgwick's Ethical Maxims," *Philosophy* 34 (1959), 217-228.
Mackie, J. L. *Ethics: Inventing Right and Wrong* (London: Penguin, 1977).
McTaggart, Ellis. "The Ethics of Henry Sidgwick," *Quarterly Review* 205 (1906), 398- 419.
Rashdall, Hastings. "Professor Sidgwick's Utilitarianism," *Mind* 10 (1885), 200-226.

[45] I wish to thank Wayne Sumner, Thomas Hurka and, especially, Robert Shaver for helpful comments on an earlier draft.

Rashdall, Hastings. *The Theory of Good and Evil*, Vol. I (Oxford: University Press, 1907).
Rashdall, Hastings. *Ethics* (London: T. C. & E. C. Jack, 1913).
Rashdall, Hastings. *Is Conscience an Emotion?* (Boston: Houghton Mifflin, 1914).
Schneewind, J. B. *Sidgwick's Ethics and Victorian Moral Philosophy* (Oxford: Clarendon Press, 1977).
Shaver, Robert. *Rational Egoism* (Cambridge: University Press, 1999).
Sidgwick, Henry. "The Establishment of Ethical First Principles," *Mind* 4 (1879), 106-111.
Skelton, Anthony. "Schultz's Sidgwick," *Utilitas* 19 (2007), 91-103.
Smith, Michael. "Does the Evaluative Supervene on the Natural?," *Well-being and Morality: Essays in Honour of James Griffin*, eds., Roger Crisp and Brad Hooker (Oxford: University Press, 2000), 91-114.
von Gizycki, Georg. "Review of *The Methods of Ethics*", *International Journal of Ethics* 1 (1890), 120-121.
Williams, Bernard. "The Point of View of the Universe: Sidgwick and the Ambitions of Ethics," *Making Sense of Humanity* (Cambridge: University Press, 1995), 153-171.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 210-229

Sidgwick on Virtue

Robert Shaver
Department of Philosophy
University of Manitoba
bshaver@cc.umanitoba.ca

ABSTRACT
Sidgwick's arguments for hedonism imply that virtue is not a good. Those arguments seemed to many wholly unpersuasive. The paper analyzes them, focusing also (especially in the final Appendix) on many changes Sidgwick made on chapter XIV of Book III through the various editions of the *Methods*. From an analysis of the first sections of this chapter, it emerges that Sidgwick employed two different argumentative schemes, one against the view that virtue is the sole good and the other against the much more diffused claim that virtue is one of the goods. These arguments can be fully understood in the context of Sidgwick's general claim that only "desiderable conscious life" is good. Sidgwick's general point is that virtue, insofar as it is valuable as an end, is so because of the feelings or consciousness associated with it.

Sidgwick's arguments for consequentialism seem, for a time, to have been wholly persuasive. Until Prichard's "Does Moral Philosophy Rest on a Mistake?" there are at best few deontologists, and even after Prichard, deontology did not revive until Carritt and Ross.

Sidgwick's arguments for hedonism seem to have been almost wholly unpersuasive.[1] Both ideal utilitarians and their deontological opponents agree that there are intrinsic goods other than pleasure. Virtue is seen not only as good, but as the most important good. This is the view of Hayward, Rashdall, Prichard, Ross, Carritt and Ewing.[2] Moore, though less enthusiastic, agrees that virtue is

[1] Hayward writes that "Sidgwick has done for [hedonism] what Plato did for his idealistic metaphysics, he has shown that the opposing arguments are almost — if not quite — as strong as the arguments in its favour" (F. H. Hayward, *The Ethical Philosophy of Sidgwick* (London: Swan Sonnenschein, 1901) p. 226).

[2] See, for example, Hayward, Philosophy ch. 8; Hastings Rashdall, *The Theory of Good and Evil* (London: Oxford University Press, 1924) v. i pp. 64-5, 71-3, 75-6, 94, 100-1, 267, *Ethics* (London: T. C. and E. C. Jack, 1913) pp. 27, 51, 64-6, 70, 72, "Professor Sidgwick's Utilitarianism," *Mind* o.s. 10, 1885; H. A. Prichard, *Moral Writings* (Oxford: Clarendon, 2002) pp. 11-12, 55-6, 61-2, 99-100 (later he claims that virtue is the only good (p. 173; Prichard to Ross,

at least one good.[3] It is, then, worth examining Sidgwick's arguments against virtue as a good, to see where, if anywhere, he went wrong.

There is another reason to look at these arguments. Their chapter (III.XIV) of the Methods — "the most important chapter" — went through many changes through different editions.[4] The result is a bit of a mess — hardly the "pure white light" for which Sidgwick is famous.[5] In the *Appendix*, I document these changes.

One preliminary: Different proponents of virtue mean slightly different things by "virtue." In *The Right and the Good* Ross thinks of virtue as the possession of certain desires, especially "the desire to do one's duty, the desire to bring into being something that is good, and the desire to give pleasure or save pain to others."[6] Prichard and Carritt separate "virtue" — desires such as the desire to help others out of sympathy, or to act courageously "from a sense of shame at being terrified," without thought of duty — and "moral goodness," the desire to do one's duty (where one thinks of the action *as* one's duty).[7] Thus Carritt defines virtuous dispositions as "those which lead people to do impulsively and effectively what reflection would generally or often show to be obligatory." Sympathy is his main example.[8] (I shall follow Ross in grouping both sorts of desire as "virtue."[9]) In *Foundations of Ethics* Ross includes, in addition to desires, acts of will, emotions such as satisfaction at the pleasure of another or sorrow at her pain, and "character," the state underlying these desires, willings, and emotions which exists even when they do not.[10] Rashdall includes one's will, desires (to do one's duty or to help others), feelings, emotions, and moral beliefs. Moore includes "a love of some intrinsically good consequence which [one] expects to

---

Bodleian Library, Oxford, Ms. Eng. Lett. d. 116, December 20, 1928)); W. D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930) pp. 134, 150-4, *Foundations of Ethics* (Oxford: Oxford University Press, 1939) pp. 275, 283-4, 290-2; E. F. Carritt, *Ethical and Political Thinking* (Oxford: Oxford University Press, 1947) pp. 66, 83, 85-6, 90; A. C. Ewing, *Ethics* (New York: Free Press, 1953) pp. 46-7, 61.

[3] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903) pp. 177-9, 217-19, *Ethics* (New York: Oxford University Press, 1965) p. 102.

[4] Hayward, *Philosophy* p. 220.

[5] Brand Blanshard, "Sidgwick, the Man" *Monist* 58, 1974, 349.

[6] Ross, *Right* p. 134.

[7] Prichard, p. 16. See also pp. 55-6, 61-2, 154, 160, 216, 218.

[8] Carritt, p. 85.

[9] Ross, *Right* p. 161.

[10] Ross, *Foundations* pp. 290-3.

produce by his action or a hatred of some intrinsically evil consequence which [one] hopes to prevent by it" and "the emotion excited by [the thought of] rightness."[11] Ewing includes willings, emotions such as love, and attitudes to others (such as displayed in fairness).[12]

In his initial discussion of virtue, Sidgwick has much the same view. Virtue is "a quality of the soul or mind."[13] It is manifested in volitions and (for some virtues, such as gratitude, benevolence, and purity) in emotions or feelings (222-3, 226). One's motive can be love of virtue or duty, or certain natural affections, such as humility or spontaneous sympathy (223, 225, 226). For some virtues, such as justice and veracity, we do not require either a thought of duty or an emotion, but rather just a "settled resolve to will" (224).

1. In III.XIV, Sidgwick begins (§ 1.) by rejecting the view that "'General Good' consists solely in general Virtue," or that "Virtue…constitute[s] Ultimate Good," or is "the sole Ultimate Good" (392, 394, 395). This " — if we mean by Virtue conformity to such prescriptions and prohibitions as make up the main part of the morality of Common Sense — would involve us in a logical circle; since we have seen that the exact determination of these prescriptions and prohibitions must depend on the definition of this…Good" (392). Sidgwick takes himself to have established that the relevant prescriptions and prohibitions are rules concerning the production and just distribution of goods. We do not know the content of the rules without knowing what is good; being told that what is good is just conformity to the rules is useless.

---

[11] Moore, *Principia* pp. 177, 179. Moore thinks the latter of small value when it lacks the hatred found in the former (218-19).

[12] Ewing, *Ethics* pp. 47, 61, 67, *Second Thoughts in Moral Philosophy* (New York: Macmillan, 1959) pp. 106, 132, 134-5, 140.

[13] Henry Sidgwick, *The Methods of Ethics* (Indianapolis: Hackett, 1981) p. 222. Subsequent parenthetical references are to the seventh edition of *The Methods*. Parenthetical references to earlier editions give the edition followed by the page. The first edition came out in 1874, the second in 1877, the third in 1884, the fourth in 1890, and the fifth in 1893 (all London: Macmillan). Other works by Sidgwick cited are *Outlines of the History of Ethics* (OHE) (London: Macmillan, 1896), "Some Fundamental Ethical Controversies" (FC), *Mind* o.s. 14, 1889, and *Lectures on the Ethics of T. H. Green, Mr. H. Spencer, and J. Martineau* (GSM) (London: Macmillan, 1902).

Against this argument — call it the circle argument — one could deny that virtue is conformity to rules. One might claim, for example, that to be virtuous is to have certain emotions, or a certain will, or certain knowledge, or a certain disposition. As long as these other things could be specified without introducing other goods, the circle argument is evaded.

Sidgwick considers most of the proposals just suggested.

(i) He admits (§ 2.) that if the virtuous person is simply one with a will to do what she takes to be right, the circle argument fails (394). But he objects, plausibly, that (a) we think some course of action *is* right, and other goods are needed to explain what makes it right, and (b) we think the will to do what one takes to be right is not always the will one ought to have, implying that there are other goods that limit it (394-5). (b) is not conclusive: one could hold that this will is not always the will one ought to have, without thinking that the reason is that other things are good. For one might think, with the deontologists, that an ought-claim such as "one ought not to will to do what one takes to be right" can be justified without relying on any claim about the good (though Prichard, Carritt and Ross themselves would not make this particular ought-claim). I consider deontology below.

(ii) Sidgwick argues that we value certain types of dispositions only because of the feelings or actions that realise them (393-4). He might be wrong here: we might value a disposition, say, to feel sorrow at the pain of others even if one never encounters anyone in pain and so never feels the sorrow. We might think a person who would feel sorrow is better than a person who would not, even if neither encounters a case that makes their difference manifest. But many will see no value in a useless disposition. And it would be odd to value *only* the disposition, and that is what would be needed to avoid the circle argument.[14]

(iii) Sidgwick does not consider explicitly the suggestion that virtue consists in having certain emotions. At times, he writes as if virtue is strictly a matter of "conduct" or "action" (395, 396). But he does mention feelings when discussing the character suggestion, and, as noted, feelings are prominent in the earlier discussion of some of the virtues. Presumably, however, he could again argue that

[14] Sidgwick might also note the current scepticism about the existence of these traits or dispositions inspired by situationist psychologists. For discussion of the literature and its upshot for virtue, see Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99, 315-31 and John M. Doris, *Lack of Character* (Cambridge: Cambridge University Press, 2002).

it would be odd to value *only* feelings. A virtuous person does not only feel sorrow at the pain of others, but acts to alleviate the pain in at least some cases. Moreover, the obvious explanation for why sorrow at pain is good is that pain is bad.

(iv) The proposal that virtue is knowledge is historically important, since it is made by the Stoics, the most prominent defenders of the view that virtue is the only good. Sidgwick objects with another application of the circle argument: if, as he suggests Plato, Aristotle, and the Stoics thought, the knowledge in question is knowledge of what is good, no guidance is forthcoming (376-7, OHE 76).[15] He notes, however, that the Stoics could reply that the knowledge is knowledge of what is to be preferred or rejected, and that what is to be preferred or rejected can be learned by observing how Nature has designed us (378n1, OHE 79-80). For example, Nature has designed us not to mutilate ourselves, hence self-mutilation is to be rejected. Against this, Sidgwick does not make the standard objection that speaking of what is to be preferred or rejected is just another way of speaking of what is good or bad. Nor does he object (though he surely would) to the theological beliefs he argues are needed to give the appeal to Nature normative force (81, OHE 77-9). Instead, he objects that the appeal to Nature's design does not provide consistent guidance: on one interpretation, it recommends rejecting what is "artificial and conventional;" on another interpretation, it recommends accepting what is established; and the Stoics did not show the superiority of one interpretation to the other (378n1, OHE 81-2). Similarly, he earlier objects that we need guidance when we have conflicting impulses, so one must be able to identify the impulses whose selection counts as conforming to Nature. But there is no way to do this: neither the impulses that are most common nor first nor independent of human action are plausible candidates for impulses that ought to be followed (81-2).[16]

Sidgwick has a further argument against taking knowledge as the good. Later in III.XIV, he will argue that knowledge is not even one good. I consider this argument below. If it works, it would also rule out the specific sort of knowledge the Stoics valued.

---

[15] For ancient versions of this objection to the Stoics and consideration of Stoic replies, see Gisela Striker, *Essays on Hellenistic Epistemology and Ethics* (Cambridge: Cambridge University Press, 1996) pp. 238-9, 217-219, 320-324.

[16] Sidgwick does not mention the Stoics here, but the problem of selecting impulses when they conflict was raised, and not solved, by the Stoics and their critics. For discussion, see Striker, pp. 219, 258-261.

A second strategy against the circle argument is to admit, as the suggestions above do not, that virtue is at least in part conformity to rules, but to argue that the rules do not need an account of the good to make them precise. For example, one might hold, with the Stoics, that the rules depend only on an account of what is to be preferred or rejected.[17] Or one might hold, with deontologists, that since at least some rules are not justified by appeals to the good, they can be made precise without such an appeal. I have discussed the Stoic suggestion: Sidgwick seems to admit that this evades the circle argument; he argues that it falters when it tries to give an account of what is to be preferred or rejected. But the deontology strategy requires comment.

Sidgwick does not consider the deontology strategy, presumably because he takes his earlier discussion of common sense morality to have shown the need for an account of the good. But even without this earlier argument, he could note that deontologists such as Ross give a role to the good. Ross subsumes four of his seven prima facie duties under the general duty of promoting the good. And the duties that Ross argues cannot be subsumed, such as the duty to keep promises, are independent of the good only in the sense that it can be right to fulfill them even when an alternative action would produce more good. Considerations of the good still enter into deciding whether one has an obligation to keep a particular promise. Thus Ross holds that whether a promise is binding depends on unspoken qualifications, and these qualifications seem to specify that keeping the promise will still bring about the good foreseen at the time of making it (or at least some good). For example, I am not bound to keep a promise to replace a string on a fiddle of someone about to die (or who no longer wants it replaced), but I am bound to keep a promise to make good the financial loss caused by my breaking a string on the dying man's fiddle, since the dying man's heirs would otherwise lose.[18] Further, how easily the prima facie duty to keep a promise can be overridden by other duties depends in part on the value of the promised service to the promisee.[19] More generally, Ross makes it a neces-

[17] Rashdall suggests this as a reply in defence of "Green's Stoicism," though Rashdall himself rejects it ("Utilitarianism" 206-8). Tom Hurka suggests it, noting Rashdall, for the Stoics (Thomas Hurka, *Virtue, Vice, and Value* (New York: Oxford University Press, 2001) p. 9). Sidgwick himself writes that "the Stoic distinction between *Good and Evil*, and *Preferred and Rejected*, is very much wanted by Green," though it is not clear that he has the circle argument in mind (GSM 99).

[18] Ross, *Foundations* pp. 94, 110; also 95-7.

[19] Ross, *Foundations* p. 100.

sary condition on the performance of duty that some good is produced.[20] Ross can say all this without failing to be a deontologist, since it remains true that sometimes one ought to produce less good rather than more; but (as I think he would admit) the good still has a role in making precise even duties that are not duties to produce the most good.

Sidgwick does mention what could be considered an example of the deontology strategy. He notes that "qualities commonly admired, such as Energy, Zeal, Self-control, Thoughtfulness, are obviously regarded as virtues only when they are directed to good ends" (392-3). One might disagree: I think we sometimes do treat these qualities as virtues even when they are directed to bad ends. Some admire these qualities in, for example, criminals. But Sidgwick could reply that our admiration is probably limited by consideration of the good — our admiration for a criminal's zeal may diminish if, say, he is a mass murderer rather than one of Ocean's Eleven.[21] And even if this is not so, it would be implausible to think that *all* virtues are independent of the good.

I conclude that the circle argument, while not as clear-cut as it might appear, can be defended with various follow-up arguments. But the biggest objection concerns its target: it hurts only those who take virtue to be the sole good. Although this might be the position of Green and (perhaps) Bradley, and Sidgwick directs the circle argument against Green, it is not the "very commonplace" position of Rashdall, Ross, etc. (GSM 73-6).[22] The view that "the character realised in and developed through Right conduct [...] is the sole Ultimate good [...] is not implied in the Intuitional view of Ethics: nor would it [...] accord with the moral common sense of modern Christian communities" (3). "[I]t is not commonly held that the whole Good of man lies in...obedience to moral rules" (391). Sidgwick realises he needs, and goes on to give, a different argument against the view that virtue is one good.

Despite this, Sidgwick is sometimes misleading. After giving the circle argument and considering replies, such as taking the virtuous person to be one who

---

[20] Ross, *Right* p. 162.

[21] Anthony Skelton noted that this is not Sidgwick's actual reply. Sidgwick thinks our admiration is "quasi-moral" and that "we certainly should not call them virtuous" (219).

[22] Rashdall, "Utilitarianism" 208. It is worth noting, however, that some seem to assume that there is but one good — this characterises the debate between Hayward and E. E. Constance Jones (Hayward, "The True Significance of Sidgwick's 'Ethics,'" "A Reply to E. E. Constance Jones," and Constance Jones, "Mr. Hayward's Evaluation of Professor Sidgwick's Ethics," *International Journal of Ethics* 11, 1900-1, 175-87, 360-5, 354-60).

simply wills to do what she takes to be right, he concludes that "reflection shows that [virtues and talents] are only valuable on account of the good or desirable conscious life in which they are or will be actualised, or which will be somehow promoted by their exercise" (395). If this is intended as a restatement of the point that types of character are valuable only because of the feelings or actions that realise them, it is unobjectionable.[23] But Sidgwick immediately starts the next section by writing that "particular virtues and talents and gifts are largely valued as means to ulterior good," as if this has been established (396). Again, if this is intended as a restatement of the point about character, it is fine — but in both places one has the impression that Sidgwick thinks he has shown something more, namely that the will, desire, and emotion involved in virtue are valuable only as means to something else. Hence the discussion on 396 concerns whether something can be valuable as both means and end; this seems to assume not just that types of character are valuable only because of the feelings or actions that realise them, but also that it has been shown that the feelings or actions are largely valued as means. The circle argument does not show that.[24]

One possibility is that Sidgwick takes the circle argument to show the presence of another good, and then that, when one looks at the other good, one sees that one finds the virtue good only when it produces this other good. The circle argument does not by itself entail that virtues are valued only as means, but it is the first step in suggesting this.[25] Thus after giving the circle argument, Sidgwick writes that

---

[23] That it is a restatement of this point is clearer in earlier editions. In the third edition, the puzzling claim from 395 is part of the paragraph rejecting dispositions (3.393). In the fourth edition, the paragraph starts with "what has been before said of Virtue regarded as a quality or element of character," where what has been before said is that dispositions are not ultimately good (4.396). The criticism of virtues and talents is that "reflection shews that they are really conceived as potentialities not valuable in themselves" (4.397).

[24] The problem may result from careless revision. The misleading claim quoted from 395 was introduced as part of a paragraph in the second edition that followed the circle argument (2.365; also 3.393). In the second edition, Sidgwick took the circle argument to show that virtue is not even one good. There it made sense to say that virtues are only instrumentally good.

[25] I owe this reading to Joyce Jenkins.

our notions of special virtues […] contain…the same reference to 'Good' […] as an ultimate standard. This appears clearly when we consider any virtue in relation to the cognate vice […] into which it tends to pass over when pushed to an extreme […].For example, Common Sense may seem to regard Liberality, Frugality, Courage, Placability, as intrinsically desirable: but when we consider their relation respectively to Profusion, Meanness, Foolhardiness, Weakness, we find that Common Sense draws the line […] by reference […] to the general notion of 'Good' (392).
My own answer to the question […] Why is the ultimate good […] held to be pleasure? is, that nothing but pleasure appears to the reflective mind to be good in itself, without reference to an ulterior end; and in particular, reflection on the notion of the most esteemed qualities of character and conduct shows that they contain an implicit reference to some other and further good (GSM 107).

Noting the reference to good shows nothing about virtue's intrinsic desirability or whether it is good in itself. But if once one sees the good, one also sees that this good determines whether the virtue is valuable, it becomes at least plausible to think that the virtue is merely a means to the good (though Sidgwick does not explicitly give an argument for the latter claim until later in the chapter).

2. In the next section (§ 3.), Sidgwick gives what Tom Hurka reads as his argument against the position that virtue is one good.[26]

Shall we then say that Ultimate Good is Good or Desirable conscious or sentient Life — of which Virtuous action is one element…? [...] [T]he fact that particular virtues…are largely valued as means to ulterior good does not necessarily prevent us from regarding their exercise as also an element of Ultimate Good: just as the fact that physical action, nutrition, and repose…are means to the maintenance of our animal life, does not prevent us from regarding them as indispensable elements of such life (396).

On Hurka's reading, Sidgwick goes on to raise an objection to this suggestion, showing a problem for thinking of physical motions in this way and then claiming that the same problem arises for virtue. It "seems difficult to conceive any

[26] Hurka, *Virtue* p. 9.

kind of activity or process as both means and end, from precisely the same point of view and in respect of precisely the same quality: and in both the cases above mentioned it is, I think, easy to distinguish the aspect in which the activities or processes in question are to be regarded as means from that in which they are to be regarded as in themselves good or desirable" (396). Physical processes are means to living, but qua physical processes have no value in themselves. What is valuable is "human Life regarded on its psychical side, or, briefly, Consciousness" (396). "In the same way, so far as we judge virtuous activity to be a part of Ultimate Good, it is, I conceive, because the consciousness attending it is judged to be in itself desirable for the virtuous agent" (397). Call this the means/end argument.

This is a puzzling argument if it is read, with Hurka, as an argument against thinking that virtue is one good.

(i) After giving the means/end argument, Sidgwick is aware that further argument is needed against the view that virtue is one good: "the Consciousness of Virtue" might still be "a part" of "Ultimate Good…conceived as Desirable Consciousness," and a part not to be identified with pleasure (398); we might take "ideal goods" such as virtue, which are not desirable merely *qua* feeling, to be a part of Ultimate Good (400); virtue is not rejected as one good until 400-1 and 402. So Sidgwick does not seem to take the means/end argument to show that virtue is not one good. Perhaps the argument is again directed against the suggestion that character, as a disposition, is intrinsically valuable. The disposition could be seen as the means to the consciousness involved in being virtuous, just as physical processes are the means to consciousness. But Sidgwick does not write of character or disposition here — he writes of "virtuous activity" or the "exercise" of virtue — and, in any case, he has already rejected character or disposition, on 393-4 (397, 396).

(ii) The problem for physical motion does not seem to be that it is valuable as a means from one point of view and valuable as an end from another point of view. Physical motion does not seem valuable as an end at all, unless Sidgwick is thinking that consciousness is constituted by (rather than caused, perhaps only in part, by) physical motions. Virtue, on the other hand, does seem to many to be valuable as a means (to pleasure, for example) and as an end, and there seems nothing incoherent about this. It is true that different properties of virtue have these values — virtue qua producer of (say) pleasure is valuable as a means; virtue qua, for example, the occurrence of certain desires, is valuable as an end. But it is unclear why this is a problem. Perhaps Sidgwick is assuming that vir-

tue is by definition a producer of some further goods, where this is intended to rule out the possibility that virtue is anything else, such as the occurrence of certain desires or feelings.[27] But (a) he has not established this definition; (b) the definition conflicts with his initial descriptions of virtue in III.II, noted above, and his treatment of many particular virtues in Book III;[28] (c) he does not need this definition to give the circle argument, since that argument turns not on a definition, but rather on the specific results of the examination of common sense morality earlier in Book III, namely that virtues such as benevolence and justice make reference to further goods to be maximised or distributed fairly (392-3); and (d) the definition does not rule out the possibility that virtues have other qualities (not true of them by definition), and these qualities could be valuable.

I think this section of III.XIV is better read, not as arguing against virtue as one good, but as arguing in favour of "desirable conscious life" as what is good. On this reading, Sidgwick first eliminates physical motions. The point of the means/end discussion is merely that once the distinction between means and end is made, there is no plausibility in thinking physical motions are valuable. In the next paragraph (396-7), Sidgwick eliminates conscious life that is not desirable. (There is no mention of means and end here.) When Sidgwick then turns to virtue, his claim is that "in the same way" we see there is no value to "virtuous activity" apart from "the consciousness attending it." In the earlier argument against character as a disposition, Sidgwick is careful to say that what might have value is actions *or* feelings.[29] Here he is eliminating actions. I do not read "in the same way" as referring back to the means/end argument, but to the more general point made against physical motion and undesirable conscious life, namely that once we view them on their own, neither is valuable. On my read-

---

[27] This is Hurka's reading (*Virtue* p. 10). He then argues that the means/end argument fails because this definition is inadequate (11).

[28] For the latter, see, for example, pp. 239, 243-5, 249, 250, 253-4, 258-60, 262, 322-4, 326, 346.

[29] A disposition "can only be defined as a tendency to act or feel in a certain way…and such a tendency appears to me clearly not valuable in itself but for the acts and feelings in which it takes effect […]. When, therefore, I say that effects on character are important, it is a summary way of saying that…the present act or feeling is a cause tending to modify importantly our acts and feelings in the indefinite future: the comparatively permanent result supposed to be produced in the mind or soul, being a tendency that will show itself in an indefinite number of particular acts and feelings, may easily be more important in relation to the ultimate end, than a single act or the transient feeling of a single moment […]".(393-4).

ing, then, Sidgwick is not arguing from a definition of virtue as instrumental to producing other goods and the view that what is by definition instrumental to producing other goods cannot plausibly be itself intrinsically good. He is instead arguing that virtue, insofar as it is valuable as an end, is so because of the feelings or consciousness associated with it.[30]

One piece of support for my reading comes from the evolution of III.XIV. Sidgwick introduced the argument that physical motions have no value in the third edition. There the argument proceeds as I suggest, as an argument by elimination for the conclusion that desirable conscious life is what is valuable: non-conscious life has no value; undesirable conscious life has no positive value. There is no mention of means and ends (3.395). When the means/end discussion is added, in the fifth edition, it is inserted into the paragraph arguing that physical motions have no value. It would be odd if this insertion were the key argument. It seems better read as merely correcting one who might think that since physical motions are "indispensable elements" of our life, they are intrinsically valuable.

A puzzle about this section remains, however. After concluding that "the consciousness attending" virtuous activity is good, Sidgwick notes that virtue also has value as a means. He then writes that "[w]e may make the distinction [presumably between virtue as means and virtue as end] clearer by considering whether Virtuous life would remain on the whole good for the virtuous agent, if we suppose it combined with extreme pain." Sidgwick thinks not. One "would hardly venture to assert that the portion of life spent by a martyr in tortures was in itself desirable" (397).

It is tempting to read Sidgwick as arguing that whatever consciousness comes with virtue is less valuable than pleasure, since we would trade that conscious-

[30] J. B. Schneewind seems to have a similar reading of the means/end argument, but he takes it to be directed against "the position that conscious virtuous action might be part of the ultimate good if we take the ultimate good to be desirable conscious life, and add that desirable conscious life has many components, of which virtuous action is one" (Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy* (Oxford: Clarendon, 1977) p. 315). This describes the position Sidgwick does not arrive at until 400, after the means/end argument, and which he attacks with arguments on 400-1 and 402, considered below. Arguing merely that it is conscious life that is valuable, as opposed to something entirely outside of consciousness, does not rule out taking virtue, insofar as it involves consciousness, to be valuable. But since the reflection argument on 400-1 is very similar to the means/end argument, Schneewind's reading is understandable.

ness to avoid pain. But Sidgwick could not expect agreement on this point; and showing it would not show that virtue is not a good (though a lesser one). It is, alternatively, tempting to read Sidgwick as arguing that it is the pleasurable consciousness attached to virtue that is valuable.[31] But since he goes on to treat the consciousness attending virtue as a live option for being good, and different from pleasure, on 398 and 400, this cannot be right either.

It is better to read Sidgwick as again concerned with "virtuous activity" in particular (397). His point is then that if virtuous activity brought no good consciousness with it (whether that consciousness be pleasure or something else), we would not value it. On this reading, when Sidgwick mentions the pain of the martyr, pain is standing in for "no good consciousness" (perhaps because extreme pain leaves one conscious of nothing else). This reinforces the point made throughout the section, that only consciousness has value, without making idle the arguments to come.

If this reading is correct, Rashdall makes a good point against Sidgwick, at least about the argument up to this point. Rashdall objects that virtue consists of desires, volitions, judgments, attentions, and emotions, and that these

> are actual elements of consciousness [....] When we pronounce character to have value, we are just as emphatically as the Hedonist pronouncing that it is in the actual consciousness that value resides, and in nothing else. It is the actual consciousness of a man who loves and wills the truly or essentially good and not mere capacities or potentialities of pleasure-production such as might be supposed to reside in a bottle of old port, which constitutes the "goodness" or "virtue" which is regarded as a "good" [...] by the school which Professor Sidgwick is criticizing [...] But for the difficulty which Sidgwick seems to make of the matter, it would have seemed unnecessary to point out that those who make "virtue" an end mean by virtue "virtuous consciousness."[32]

Through this point in the chapter, although his arguments may be successful, Sidgwick has not engaged with his real opponent.

---

[31] For this reading, see Schneewind, p. 316.

[32] Rashdall, *Theory* i pp. 64, 65; also, directed at the third edition, "Utilitarianism" 224. Hayward makes the same point: "Sidgwick's argument is sound but unnecessary" (*Philosophy* pp. 221; 199, 222, 230-1).

3. In the next section (§ 4.), Sidgwick clarifies the position of this opponent. To value knowledge is to value something that is within consciousness (a belief) and something which goes beyond consciousness (that the belief is true, justified, etc.). To value virtue is to value something that is within consciousness (volitions, etc.) and something which goes beyond consciousness (that the volitions, etc. are morally good). The arguments above do not conclude that what is valuable must be confined to consciousness. They conclude only that consciousness must be part of what is valuable. Hence they do not rule out knowledge or virtue as valuable.

In the following section (§ 5.), Sidgwick argues that virtue, so understood, lacks value. The first argument is that he finds it "clear after reflection that these objective relations of the conscious subject, when distinguished from the consciousness accompanying and resulting from them, are not ultimately and intrinsically desirable; any more than material or other objects are, when considered apart from any relation to conscious existence" (400-1). Call this the reflection argument.

The reflection argument asks why adding something (a material object) that has no effect on consciousness adds no value, whereas adding something else (the truth or justification of a belief or the goodness of a volition) that has no effect on consciousness adds value. Defenders of knowledge or virtue can reply that the whole formed by a belief and its truth or justification has more value than the belief has on its own, or that the whole formed by a volition and its goodness has more value than the volition has on its own.[33] The difference between these additions and adding a material object is just that no whole with greater value is brought about by adding a material object.

Sidgwick might avoid this reply by taking a certain lesson from the discussion of material objects. The reason for rejecting material objects is that consciousness is unchanged; that is a reason for rejecting knowledge and virtue as well.[34] The worry is that lacking effects on consciousness may be conclusive against material objects, since we cannot see how else they could affect value, but not

[33] I put the point in terms of the "holistic" rather than "variability" view, but some might want to put it in terms of the latter — rather than thinking a whole with more value is formed, one might say that the belief or volition itself acquires more value. For the distinction, see, for example, Thomas Hurka, "Moore in the Middle," *Ethics* 113, 2003, 606-7.

[34] I owe this suggestion to Joyce Jenkins.

against other things, such as having correct and justified belief or virtue, that could affect value in other ways.

The second argument is an appeal to "a comprehensive comparison of the ordinary judgments of mankind" (400). There are cases "in which the concentration of effort on the cultivation of virtue has seemed to have effects adverse to general happiness, through being intensified to the point of moral fanaticism." In such cases, "we shall…generally admit that…conduciveness to general happiness should be the criterion for deciding how far the cultivation of Virtue should be carried" (402).[35] Call this the criterion argument.

One problem with the criterion argument is that it fails to show that virtue is not a good. Even if we prefer happiness to virtue, we might think, with Ross, that of two worlds equal in happiness but unequal in virtue, the world with more virtue is better.[36]

Another problem is that some may disagree. Consider not a moral fanatic, but rather one with so much sympathy that she sometimes helps others when she is not qualified to do so, and so makes matters worse. Some may think a world with such people, and less happiness, is better than a world with less sympathy and more happiness.[37] (Others, again, condemn those who "mean well.")

---

[35] Similarly, Sidgwick writes that "when Virtue and Happiness are hypothetically presented as alternatives, from a universal point of view, I have no doubt that I morally prefer the latter; I should not think it right to aim at making my fellow-creatures more moral, if I distinctly foresaw that as a consequence of this they would become less happy. I should even make a similar choice as regards my own future virtue, supposing it presented as an alternative to results more conducive to the General Happiness" (FC 487). Rashdall replies by inviting "the reader to say whether he can accept [this] as a correct representation of his own moral consciousness — or of Henry Sidgwick's" (*Theory* i, p. 70).

[36] Ross, *Right* p. 134.

[37] Rashdall gives the following (now unconvincing) example: "On what other grounds can we either explain or justify [Common Sense's] emphatic condemnation of suicide in cases where it is clearly conducive to the happiness of the individual and of all connected with him?" ("Utilitarianism" 219). Elsewhere he gives examples of bullfighting, Roman wild-beast and gladiatorial fights, German students' face-slashing duels, coursing, pigeon-shooting, and drunkenness ("and we should think a man's conduct in getting drunk worse instead of better if he had carefully taken precautions which would prevent the possibility of his doing mischief…while under the influence of his premeditated debauch") as worse than their absences even if they maximise pleasure (*Theory* i pp. 97-9). (Rashdall also argues, more ambitiously, that Sidgwick is inconsistent in taking pleasure but not virtue as good. I examine this argument in *"Utilitarianism,"* forthcoming in the *Oxford Handbook of the History of Ethics*, ed. Roger Crisp).

I should note an alternative interpretation. J. B. Schneewind suggests that Sidgwick's point on 402, and earlier when rejecting the good will, is that "there is a limit to the extent to which we think it supremely good to act according to one's moral convictions: and the limit is determined by the utilitarian principle. Here a dependence argument shows that the good will [or virtue more generally] cannot be an ultimate good, for its limits are determined by the claims of another good and its own directives may be overridden in the name of that other good."[38]

The problem is that Sidgwick seems to understand by an "ultimate" or "intrinsic" good a good which is good as an end rather than just as a means. He does not seem to mean a good that is not limited by other goods. For example, when he considers whether virtue could be both a "means to ulterior good" and "an element of Ultimate Good," he comments that "it seems difficult to conceive of any kind of activity or process as both means and end…and…it is…easy to distinguish the aspect in which the activities…are to be regarded as means from that in which they are to be regarded as in themselves good or desirable" (396). Sidgwick takes his opponents to hold that virtue (and other things) "are ends independently of the pleasure derived from them" (401). He explains that he means by "'Ultimate Good' […] that which is Good or Desirable *per se*, and not as a means to some further end" (407n). If so, showing that one good is limited by another does not show that the limited good is not ultimate. It might be merely a lesser ultimate good. Thinking that we always trade virtue for happiness, even if true, does not show that the value of virtue is dependent on its production of happiness. Schneewind's interpretation has the advantage of making the criterion argument valid. But it has the disadvantage of not fitting Sidgwick's claims about means and ends and, more importantly, makes Sidgwick talk past those, like Ross, who hold the usual understanding of "intrinsic" or "ultimate."

Sidgwick might have tried a different strategy. In the case of knowledge, the appeal to the ordinary judgments of mankind proceeds by noting that knowledge is valued in proportion to the happiness it brings. The connection to happiness can explain why we value even apparently fruitless knowledge, both because we know that such knowledge can "become unexpectedly fruitful" and

[38] Schneewind, p. 314. This interpretation also fits the passage concerning Liberality, etc., quoted earlier.

because its pursuit is itself pleasurable and shows a disposition likely to produce fruitful knowledge (401). Call this the proportion argument.

It does not follow that knowledge is good only as a means. First, each piece of knowledge might have the same value, with the proportion claim made true by combining these values with the differing amounts of happiness produced.[39] Second, pieces of knowledge might vary in value, but not so much that less valuable knowledge that produces greater happiness is ranked lower than more valuable knowledge that produces less happiness. Since the proportionality claim is hardly precise, it might be hard to discount this possibility. Third, Ross's example of two worlds equal in happiness and unequal in knowledge convinces some that the proportionality claim is false.[40]

Sidgwick might have tried a parallel argument for virtue. He does not explicitly say that virtues are valued in proportion to the happiness they bring. But he makes the similar claim that utilitarianism explains our ranking of duties (425-6); it is plausible to think that the "minor" virtues — Sidgwick lists caution, decision, good humour, meekness, mildness, gentleness, placability, mercy, liberality, politeness, and courtesy (236, 253, 321, 324-5) — are minor because they are usually less productive of happiness than virtues such as benevolence and justice; he argues that virtues such as purity, courage and humility, which seem to be admired independently of happiness, really do, insofar as they are admired, contribute to happiness (or other virtues) (332, 334, 355, 356n, 429, 450-3, 456); and when Sidgwick later writes that in III.XIV he "tried to show that Common Sense is unconsciously utilitarian in its practical determination of those very elements in the notion of Ultimate Good or Wellbeing which at first sight least admit of a hedonistic interpretation," he suggests that the utilitarian ranking of pieces of knowledge is to be paralleled by a utilitarian ranking of virtues (453-4).

Just as with the proportionality argument against knowledge, this is hardly conclusive. But two of the reasons for resisting the proportionality argument in the knowledge case seem less telling in the case of virtue. Few proponents of virtue would claim that each instance of virtue has the same value. And perhaps, given the importance ascribed to virtue, it would be difficult to hold that differences in the values of virtues are sufficiently small as to exclude the possibility

[39] Skelton noted that Ross himself rejects this; see *Right* p. 139.
[40] Ross, *Right* p. 139.

that a more important virtue that produces less happiness has more value than a less important virtue that produces more happiness.

Ross's appeal to worlds of equal happiness and unequal virtue remains. But here Sidgwick might note that Ross's verdict is controversial — in my experience, at best half agree with Ross — and, as with knowledge, Ross's intuition might be explained away, given how difficult it is to imagine something normally so useful as making no difference. Slight variations in the presentation of the case also seem to hurt Ross. For example, say I could increase virtue by writing an inspirational book in moral philosophy, but this would make no difference to the amount or distribution of happiness in the world. (Say the book increases the number of actions done out of duty, but that in all these cases self-interest would have led to the same action.) Many think it does not matter whether I write the book or not.

This, at any rate, seems the sort of argument Sidgwick should have stressed. If he had, the supporters of virtue who came later would at least have had to work harder. It is regrettable that so much of III.XIV is spent on other matters.

Appendix

III.XIV in the final edition incorporates, not always smoothly, changes made over the first four editions. It may be helpful to briefly chart these changes.

In the first edition, Sidgwick notes that "the majority of moral persons would probably declare that Virtue is the *chief* good [but] very few would maintain that the *only* thing in life intrinsically desirable is the habit of obeying moral rules" (1.369). Against the view that virtue is the only good, he gives a quick version of the circle argument (1.369, 376). There is also a version of the argument against dispositions as being of value, but Sidgwick does not take this to count against virtue, but rather to specify that virtue is a matter of "conscious action and feelings" (1.369). Virtue is rejected as a good, along with other objective relations such as knowledge, on the basis of the reflection argument (1.371-2). The criterion argument against virtue and the proportion argument against knowledge do not appear.

In the second edition, the circle argument is expanded to roughly its final form (2.364-5). Virtue is dismissed on the basis of it: it follows from the circle argument that "we cannot, without manifest divergence from Common Sense, introduce [virtue] in a scientific explanation of the nature of Ultimate Good"

(2.365). This is a blunder not made in the first edition, and not wholly corrected until the fifth: the circle argument excludes virtue only as the sole good, not as one good. The prominence of the circle argument, and the focus on virtue as the sole good, may be due to Bradley's *Ethical Studies* and *Professor Sidgwick's Hedonism*, both of which appeared between the first and second editions. Bradley sometimes claims that the sole good is "function," and sometimes treats "function" and "virtue" as interchangeable.[41]

In the second edition, knowledge is presented as an alternative to virtue or happiness and is rejected by the reflection argument and the proportion argument (2.366-9).[42]

In the third edition, Sidgwick expands the argument against dispositions to roughly its final state (3.393). One difference is that he takes it to be a further argument ruling out virtue, and not just dispositions, as one good (3.394). Unlike the circle argument, this argument could show that virtue is not one good, provided one thought of virtue just as a disposition — but as Rashdall notes, defenders of virtue need not think this. This may explain why, by the fourth edition, the argument is taken to discredit only dispositions, and not virtue in general.[43]

The third edition also adds the arguments against physical processes and mere survival. They are introduced to limit what is valuable to conscious life (3.395).

The fourth edition adds the argument against the will to do what one takes to be right (4.394). Sidgwick recognises that, even if this argument succeeds, the good will could still be one good: it would be a paradox to "affirm [subjective rightness of will] to be the sole Ultimate Good" but not "paradoxical to regard the settled will to realise our duty as an essential part of ultimate good: while at the same time recognising that there are effects of right volition [...] which are also in themselves good" (4.394).[44] He then objects that if I "suppose that the

---

[41] See, for example, F H. Bradley, *Professor Sidgwick's Hedonism*, in Bradley, *Collected Essays* (Oxford: Oxford University Press, 1925) pp. 95, 96n, 97, 98, *Ethical Studies* (Oxford: Clarendon, 1927) pp. 136, 137, 138; *Hedonism* pp. 97, 98, *Studies* pp. 140-1.

[42] The second edition is very similar to "Hedonism and Ultimate Good," *Mind* o.s. 2, 1877, published in the same year.

[43] The upshot in the third edition is that "virtues or talents, faculties, habits or dispositions of any kind" are not goods (3.393). In the fourth and later editions, the upshot is that "faculties, habits, or dispositions of any kind" are not goods (4.393, 393).

[44] This is noted by Schneewind, pp. 313-14.

effects of a man's acting in accordance with his conception of what is right will be on the whole bad — according to an estimate of badness framed without taking into account the subjective rightness of the volition — , I find that this consideration of them appears to me finally decisive of their badness. In my view, therefore, this Subjective rightness of volition is not Good in itself, but only as a means" (4.395). There is no further discussion of virtue as one good. In effect, Sidgwick runs the criterion argument not against virtue in general, but against one account of virtue, as the good will.

The fifth edition (which for III.XIV is the same as the later editions) takes seriously the concession made regarding the good will: not only the good will, but also virtue more generally, has not been excluded as one good by the circle argument. Sidgwick then restores virtue to the place it had in the first edition, as an objective relation like knowledge. (Virtue has this place in the *Lectures* as well (GSM 126).) It is rejected by the reflection argument, as in the first edition, and by the new criterion argument, which generalises the point made against the good will in the fourth edition.[45]

[45] Thanks to Darcie Fehler, Adam Muller, Emily Muller, Jeff Verman, Sandy Vettese and Andrew Webb for discussion of some of the examples; to Tom Hurka for discussion of many of the moves in the paper; and to Joyce Jenkins and Anthony Skelton for detailed comments on an earlier draft.

SYMPOSIUM:
WALTER BLOCK, *Labor Economics from a Free Market Perspective*

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 232-235

On Block's Labor Economics

David Gordon
Ludwig von Mises Institute
dgordon@mises.org

Labor Economics From a Free Market Perspective[1] contains 29 essays by Walter Block. If I am not mistaken, seven further volumes of his papers are to appear. He is astonishingly prolific, and he is also well known for the large numbers of co-authors whom he has enlisted as collaborators. The present collection includes eleven co-authored papers, written with twelve different authors. The volume deals with a topic of major importance. Walter Block tells us that labor "accounts for some 70-75% of the GDP." (xix) If so, it is vital for the economist to explain how wage rates are determined. For Block, the answer admits of no doubt. Wages on the free market are determined by the marginal productivity of the workers. Suppose a firm employs ten workers to perform the same sort of labor. Each will then receive approximately what the tenth worker adds to the product, i.e., each worker will receive the marginal revenue product.

Why is this so? Employers will not pay more than this, since it would not be profitable for them to do so. If an employer, tried to pay a lesser amount, competing employers would find it profitable to outbid the low payer; they would do so until the wage approached marginal productivity.[2]

Block shows himself alert to refinements of this picture. What the worker receives is not, strictly speaking, the marginal revenue product: it is the discounted marginal revenue product. Time preference accounts for the discount: The employer normally pays the worker immediately but must wait until the product is sold before he gets money himself. Because people prefer present goods to future goods, the employer gets a premium for waiting: equivalently, the employees suffer a loss because they do not wait. This is of course the Austrian, as opposed to the neoclassical view; and Block skillfully argues that time preference is a universal feature of action. "The fact that we choose to act in the present, when we could have waited, shows that we prefer goods, the sooner the

---

[1] Walter Block, *Labor Economics From a Free Market Perspective: Employing the Unemployable* (Singapore: World Scientific, 2008). All references to this book will be by page numbers in parentheses in the text.

[2] "More technically, below the alternative cost of MRP, namely the MRP that would obtain in the next best alternative to present employment." (p.37, note 3)

better. . .By acting in the immediate future, instead of waiting for the more distant future, we also show ourselves as present oriented.' (p.39)

Block is characteristically aware of objections, and he always has a response. To the contention that wages are determined by bargaining power, Block answers that this cannot be taken as an explanatory ultimate. If wages are below the DMRP, then workers have more bargaining power; if wages are above this rate, then employers have more bargaining power; and if wages equal the DMRP, then neither side has greater bargaining power. Bargaining power drops out of the explanation: in Wittgenstein's phrase, it is a wheel on the machine that does no work. Many people who are not economists find the marginal productivity theory hard to grasp; and in a exchange with Boyd Blundell, a religion professor at Block's own university, Loyola at New Orleans, who insists on bargaining power as an independent force, Block patiently explains his position. ""Prof. Blundell maintains that worker 'productivity is virtually irrelevant' to the setting of labor's compensation. Rather, it is driven by 'bargaining power.' But the latter depends almost entirely on the former." (p.115)

Block must overcome another objection. The process by which wages below the DMRP rise depends on competing firms. Only if a rival firm exists will there be a chance for lower wages to be bid up. What happens if there is only one buyer of labor services, i.e., a monopsony exists? Block responds this situation is most unlikely to arise in the free market. "Even on the heroic assumption that monopsony is itself a logically coherent analytic construct. . .outsiders will enter the market to take advantage of the profits earned by the monopsonist; in the absence of entry barriers, monopsony, even if it could be established in the first instance, cannot long endure." (p.150) He himself rejects the entire concept, following the classic discussion of monopoly by Murray Rothbard in Chapter 10 of *Man, Economy, and State*.[3]

If Block is right, wages cannot be increased beyond the DMRP. Efforts to push wages higher will generate unemployment, since employers will not be willing to lose money by paying someone more than he is able to contribute to the product. Labor unions have as the principal purpose to force wages above the market level, and Block has little use for them. He does not deny that workers have a perfect right to form associations and to quit a job in concert. He rejects the view of W.H. Hutt that it is inherently collusive to do so. "But this [the po-

[3] Murray N. Rothbard, *Man, Economy, and State* (Auburn: Ludwig von Mises Institute, 2004), Chapter 10.

sition that collusive actions by unions exploit the community] only shows that there is all the world of difference between economists who favor a system of laissez-faire capitalism, on the one hand, and those who favor a system of national or state capitalism on the other." (p.71) But beyond this, workers' associations have no right forcibly to impede others from engaging in business with a firm that the workers wish to bring to heel. In particular, they cannot legitimately interfere with customers' accesses to the business by picketing or use force against workers whom the employer hires to replace them. These workers, Block maintains, should not be stigmatized as "scabs". Also to be deplored are laws that compel employers to deal with unions.

In practice, Block thinks, all unions engage in such wrongful activities. He does not deny the possibility of a union that acted in entire accord with freedom of contract but professes never to have found one. Accordingly, he finds nothing amiss with "yellow dog" contracts that require non-membership in a union as a condition of employment. "The Yellow Dog Contract, in addition to safeguarding employer and employee rights of free association, also serves as a remedy against union inflicted economic disarray and violence against innocent people and their property. Long live the Yellow Dog Contract. Bring it back. Now." (p.110)

In their efforts to raise wages above the free market, unions also support minimum wage laws; and legislation of this type arouse our author's well-justified ire. Minimum wage laws hurt the poor and unskilled. The laws make it unprofitable to hire, or to continue in employment, workers whose DMRP is below the minimum wage.

Why do unions, whose members normally earn well above the minimum wage, support these laws? They do so to restrict competition. Faced by high wage demands from unions, employers will be tempted to hire lower skilled workers to replace the union members, even if they have to increase the number of people on their payroll to get the job done. Minimum wage laws hinder their ability to do so.

The great majority of economists agree with Block that minimum wage laws cause unemployment. Unfortunately, a number do not. In particular, a petition signed by 350 economists, including such luminaries as Kenneth Arrow and Joseph Stiglitz, claimed that minimum wage legislation was a good idea in present conditions. But what about unemployment? These economists do not deny that sufficiently high minimum wages would cause unemployment — imagine, e.g., a

minimum wage set at $10,000 per hour. They claim, though, that that if the rate is moderate, the law will do no harm and may do some good.

Block is outraged. If the argument that minimum wage laws cause unemployment is correct, then even a "moderate" rate will result in unemployment, so long as the rate is above the DMRP of some workers. Block thinks it is a disgrace that these economists have ignored elementary principles, and he reprints the entire list of signers to call attention to their misdeed. "One of these days justice will prevail, and the eminent reputations of all those who signed the document will be called into question." (p.160)

Block must here face an objection; and, as usual, he has an effective answer. "Your theory is all well and good", the objector might say, "but careful empirical studies show that minimum wage rates do not have the dire effects you claim. What of Card and Krueger?" This study compared employment in fast food restaurants in New Jersey, which enacted a minimum wage law, with Pennsylvania, which did not. The student found no significant employment effects resulting from the law.

Block responds with a detailed criticism of their often-cited study. It is based on an inadequate sample; it suffers from other statistical failings; and it ignores the effects of earlier federally imposed minimum wages. Though best known as an Austrian economist, Block received his training in neoclassical economics and is thoroughly familiar with econometrics. His criticism of Card and Krueger illustrates what he regards as a fundamental point. The theorems of economics are established through deductive reasoning, starting from the axiom of action. (This is of course the view of Mises in *Human Action*.) As such, they cannot be refuted through empirical tests. If a test result goes counter to an established theorem of praxeology, there must be a mistake somewhere. "Even more daunting is that fact that their [Card and Krueger's] findings are contrary to economic law. . .On the level of pure theory, then, it must count against CK that — apart from the economically dubious monopsony argument — they felt no need to account for their anomalous findings." (p.150)

The collection includes papers from a wide variety of other topics as well, including immigration and reparations; but I have concentrated on a central theme. In all the papers, Blocks writes from a firm commitment to libertarianism; and he displays a complete mastery of technical economics. It is a powerful combination.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 236-247

Unblocking a Free Market Perspective in Labor Economics

Per Bylund
University of Missouri
Per.Bylund@mizzou.edu

0. *Introduction*

In *Labor Economics from a Free Market Perspective: Employing the Unemployable* (2008) Walter Block presents a seemingly comprehensive free market perspective on the economics of the labor market. From this perspective Block discusses the economics of wage determination and the minimum wage's effects on the labor market; the economic impact of labor unions and unionized regulation as well as the economics of unemployment insurance and academic tenure; and he touches on immigration, redistributive justice, and slavery reparations. There should be no surprise that the free market perspective allows Block to argue that regulation and intervention in the market cause imbalances and disequilibria and are therefore economically inefficient and undesireable. But Block goes one step further and argues that all kinds of regulation or tampering with a free market setting for voluntary interaction of individuals are simply wrong. There is no doubt that Labor Economics from a Free Market Perspective is a very provocative book.

On the one hand, it is a treatise on labor economics covering basic economic truths such as the determination of wages and the effects in the labor market of enforced minimum wage laws and unionism. Just like in Block's 1976 book *Defending the Undefendable*, from which the sub title "Employing the Unemployable" seems to be borrowed, the author investigates well-known institutions and offers thought-provoking arguments based on distinctly economic reasoning. The difference is that Block does not pick heavily disliked social phenomena to which he offers strong arguments *in favor*, which is the case in Defending the Undefendable, but argues fiercely *against* generally accepted and commonly advocated political solutions to perceived market problems. Even though many of the arguments are true from a mainstream economics point of view, most of the illustrative examples and analogies to complement them are, in a typically Blockian manner, very outspoken, shocking, and – sometimes – even infuriating.

On the other hand, *Labor Economics from a Free Market Perspective* is far from a neutral and *wertfrei* theoretical study of economic phenomena accompanied by solutions based on pure economic reasoning. It is a treatise with a distinct ideological base: it is hard-core libertarian, a view that, in its intrepid advocacy for unbridled individual liberty, itself should prove provocative to most people. This radical perspective literally permeates the book's chapters and, combined with Block's obvious fancy for taking coat-trailing standpoints, it leaves no reader unperturbed.

The author does not try to hide the fact that the book takes a clear value-based position. Contrarily, in the introduction Block explicitly states that it is "an ideological book" but that this, to the author, does not mean the approach is unscientific but only that it "takes a position on ideas" (p. xix). The position is explicit already in the title of the book and is further stressed in the introduction, where the author declares that the book "look[s] at numerous labor market issues from a vantage point of free enterprise or libertarianism" and that a "cure" to the problems discussed is available through "private property rights, the non-aggression principle and the law of free association" (p. xix).

An opponent to libertarianism and free markets would, as would economists and other representatives for the "positive sciences," find plenty of reasons to criticize Block. It is safe to say that the bulk of such criticism would target the ideologically based perspective the author has chosen, and that such critique would be based on the seemingly obvious contradiction between science and ideology. In this paper, however, I will argue that the "obvious" contradiction need not be contradictory at all. From a radically libertarian point of view it can be argued that the "is-ought problem" is partially solved, or at least inapplicable, and therefore that criticism based on "Hume's Law" may be misdirected. This is not to say, however, that Block cannot be criticized for the assumed ideological perspective on which his economic arguments are supposedly based.

In the remainder of this paper, I continue to analyze the essence of Block's argument from what I suspect is a somewhat unexpected angle: I criticize the scope of his arguments, and especially his conclusions and underlying assumptions and reasons, adopting the radically libertarian or free-market point of view – the very same view championed and utilized by Block.

1. *Hume's Law and the Libertarian Idea*

David Hume (1739-1740) famously identified that there is a significant and important difference between descriptive ("is") and prescriptive ("ought")

statements. He advised against deriving an "ought" from an "is," i.e., to draw normative conclusions based on empirical facts, without clearly explaining exactly how employed ought-statements follow from is-statements. Hume is often assigned the position that there is no solution to the "is-ought problem," and this so-called *Hume's Law* (that "ought" cannot be derived from "is") is often used to clearly distinguish between and separate positive (empirical) science from normative (ethics).

Block's "ideological book" seems to clearly violate Hume's Law in that it argues from a point of view of libertarianism (i.e., libertarian *ethics*) using primarily arguments from a distinctly positive science, namely economics. As has already been noted, however, this may not necessarily be a correct interpretation of Block's position and arguments. There are two reasons for this: firstly, Block is an economist in the Austrian tradition (see e.g. Block, 1999), which means he bases the argument on praxeological reasoning (Mises, 1949; Rothbard, 1962) rather than "mainstream" economic techniques; and secondly, libertarianism as an ideology is often portrayed as an open-ended and tolerant "system" that sets a non-restrictive *formal framework* but refuses to provide a social blueprint, which makes it less normative than most ideologies. As we shall see, free markets and libertarianism may not be different in substance or nature but only in approach or perspective.

The free market is the starting point in economic analysis and very often the "ideal" in terms of market efficiency. Market equilibrium theory, which is a cornerstone in mainstream economic analysis, shows maximum resource utilization in terms of production (supply) and, as a consequence, satisfaction of consumer wants (demand). In other words, economic theory strives to find the most efficient means to certain ends given an explicit amount of resources (inputs) and specific production functions (technologies) in a particular market. In this sense, therefore, economics as a scientific discipline is founded on a utilitarian philosophy of what is universally good (efficiency) and can therefore make claims as to what is a "better"[1] solution, even though the

[1] "Better" should here be interpreted in the strictly economic sense, i.e. "more efficient [use of resources]." The science of economics is based on the seemingly utilitarian idea that "efficient" is better than "inefficient" because of the greater possibility of satisfaction of consumer wants/demand, and therefore that increases in utility are strictly better than decreases. See e.g. the economics concept "Pareto improvement," which describes a change in which no individual is affected negatively in terms of utility and at least one individual is "better off" (Pareto, 1971). In this sense, economics claims not only to explain and predict economic phenomena but to provide a [normative] basis for decision-making.

explicit questions it tries to answer and the phenomena it tries to describe and explain are more scientific in the positive sense.

Economics in general, and especially the Austrian tradition, is deductive in nature and as such guided by a certain set of assumptions of e.g. the rationality of economic actors. Austrian economics consists of a complex set of detailed economic truths derived from the "action axiom," which states that humans take conscious action toward chosen goals. It is *wertfrei* in that it explains the functions and workings of the market and its institutions, and attempts to explain effects of certain changes in and to the market, such as entrepreneurship and production choices in the first sense and regulations and taxes in the latter.

Just like most other approaches to economics, the Austrian school does not propose or advocate an "ideal" setting or structure for the economic system. On the contrary, it relies significantly less on equilibrium analysis in its study of the market than e.g. mainstream economics. However, Austrian economics rejects statistical methods and empirical studies as means to learn about economic truths and hence adopts a purely deductive approach. As such, it does not refine or change its explanations and theories to "fit" empirical data (as is the case in "semi-deductive" mainstream economic research) and therefore it tends to maintain the truth of fundamental economic theory: that all interventions are necessarily and without exception regarded as causes of inefficiencies or distortions imposed on the economy.[2] True to form, Block's arguments throughout the book are directly aimed at these causes of inefficiencies: intervention in and regulation of the free market.

It follows from the statement above that the only state of the economy without distorted outcomes and inefficiencies is a market free from interference. Thus, even though Austrian economics is not in itself normative, it clearly shows the strictly negative effects of interventions in the market place, which inevitably provides individuals of certain moral convictions the arguments and moral reasons to espouse an unrestricted market process. There is therefore, to a certain degree, a possible link between the purely scientific study of the market/economy and the normative advocacy of unregulated/free markets.

The normative view supporting free markets as well as, or perhaps primarily, free people is often denoted libertarianism or libertarian ethics. As we have seen, libertarianism is the explicit "vantage point" of Block's study of labor

---

[2] Any and all restrictions of the market process can be shown to cause e.g. inefficiencies through discouraging profitable investments or encouraging "too risky" investments.

economics, which makes its definition and implications highly relevant to our discussion on a possible violation of Hume's Law. In Block's own words (1994:117, emphasis in original):

"Libertarianism is a political philosophy. It [is] concerned *solely* with the proper use of force. Its core premise is that it should be illegal to threaten or initiate violence against a person or his property without his permission; force is justified only in defense or retaliation."

This is clearly a prescriptive definition of the political philosophy libertarianism, and as such it should violate the aforementioned law. But this is not necessarily the case. We have already seen that economics as a science and especially its use of the "free market" equilibrium is not perfectly descriptive but includes prescriptive elements; more specifically, economics is the study of the economy using the free market as benchmark. Since libertarian philosophy, using Block's definition above, is concerned *only* with "the proper use of force" we need only investigate whether this "libertarian law" is compatible with the scientific study of the market.

Any ideology is by definition normative and therefore so is libertarianism. However, as was previously mentioned, it is much less so than competing ideologies in that it insists on a "non aggression principle" as a necessary and sufficient condition of liberty but does not predict nor prescribe the nature of liberty. The principle itself states only that "everyone may act precisely as he pleases, provided, only, that he does not initiate violence against non aggressors"[3] (Block, 2008:xix), and is therefore a definition of the necessary limits of freedom (cf. the Hobbesian state of nature in which no such legal limits to freedom exists).

As "libertarian law" states only that people are free to act and associate as they choose for as long as they do not initiate the use of physical force, a number of ideological utopias should be obtainable *within* that framework and therefore, in a weak sense, compatible with libertarianism. In other words, libertarianism is not exclusive in the sense that it excludes other-than-libertarian ways of life or organization (cf. Nozick, 1974), and therefore does not make claims for how people should lead their lives. It only limits

[3] The non-aggression principle is not exclusively libertarian, but is an important part of the so-called "natural law" tradition in which it can be traced back to St. Thomas Aquinas or even Epicurus.

individuals' actions to anything that does not do direct harm to other individuals.

This is very similar to the definition of the free market, where nothing supposedly restricts the competitive market process from bringing the quantity demanded by consumers and the quantity supplied by producers into equilibrium. A market where force exists, i.e. where contracts are breached and property rights violated, would rarely be denoted "free" – the use of force is not compatible with the voluntary exchange of goods and services. As we have already noted, economics, and especially Austrian economics, studies and predicts the negative effects of interventions in the market place. Interventions are any forceful changes to or restrictions of economic actors' behavior, which necessarily includes the initiation of physical force. The free market, therefore, is fundamentally based on a principle that is very similar, or even identical, to libertarian law. Also, libertarianism, based on the non aggression principle, cannot espouse any other economic "system" than a free market economy – all alternative ways of economic organizing would necessarily violate the fundamental principle. The free market and libertarianism are therefore, in substance, two sides of a coin: one cannot exist without [a version of] the other.

But this does not imply that the free market *is* libertarianism, or vice versa, even though they share fundamental properties both in theory and "practice." The difference lies not in substance, but in use and perspective of the concept. The former is a description of the unrestrained market whereas the latter is an image of a potential "good" society where nothing is allowed to restrict the market (in a broad sense, i.e. including basically any human interaction). Both concepts are therefore identical to the extent that they describe a state of unrestricted voluntary interaction, but different in underlying purpose. It is therefore wrong to claim that the free market is libertarian, whereas it would be correct to claim that libertarianism champions and includes the free market. The perspective and purpose, therefore, while not substance, of the free market concept is primarily positive and scientific, whereas for libertarianism it is normative.

Block is hence correct in that adopting a libertarian point of view in the study of the labor market does not compromise the scientific nature of the economic argument. But only to the extent that his libertarianism does not affect economic conclusions or distort facts and arguments through applying a distinctly libertarian perspective where such is inapplicable; the free market is an economic model of "ideal" (optimal/maximum) efficiency, but libertarianism is not. In other words, the study's scientific value is limited to the extent that the libertarianism in Block's argument is strictly the use of

the non aggression principle to explain, describe and define the free market. Hume's Law is violated only when and where Block's libertarian views are used in an explicitly normative manner, i.e. when libertarianism is used specifically as a libertarian *ethics* in addition to the voluntary nature of the free market.

It should therefore be concluded that an "ideological book" such as Block's should not automatically be dismissed as unscientific due to its ideological vantage point. Even though the vocabulary of choice in many of the articles is clearly libertarian, Block generally manages to stay on the right side of the road from a science point of view. His arguments are economically correct and straightforwardly presented; there is no obvious flaw in the logic and he gives the reader no reason to doubt the validity of the argument; economists would find it difficult to criticize Block's strictly economic reasoning. They would first, however, have to see through Block's provocative language and somewhat unorthodox examples.

2. *An Ideological Analysis*

But from a libertarian *ideological* point of view the author is not as safe from criticism. From a radical libertarian perspective the issues discussed by Block are both interesting and important, but the depth of the analysis of such intervention in the labor market is insufficient – the analysis is too limited and does not take into account all major effects of market regulation. I will here use section II in the book ("Unions") as an example, but the same line of reasoning is applicable on most arguments put forth in the book.

The starting point for Block's analysis is, as has already been discussed at length, the non aggression axiom as a distinguishing property of the free market. In Block's interpretation, the analysis is primarily from a point of view of *freedom*, a concept which has a distinctly normative flavor. By freedom Block means the rule of libertarian law and therefore non-violation of the non aggression principle – the absence of violence and coercion.

We have already shown that the concept of freedom as a distinctly libertarian ethics is necessarily normative, and that Block's claim to do the analysis from the perspective of freedom would therefore violate Hume's Law. But we have also concluded that the analysis is not in violation of the aforementioned law. The reason for this is that Block does not predominantly provide arguments from a point of view of freedom – despite his claim to do so – but provides arguments distinctly targeted at violent action as interference in the market place. Libertarian freedom might be Block's

underlying purpose and ideal, but the arguments are above all against certain instances of violent intervention in the market and not pro libertarianism *per se*. In the analysis of labor unions this fact is made explicit when that which is analyzed is only the coercive aspect of unionism – "we are defining unions as organizations that use coercive force" (Block, 2008:62) – while the non-coercive aspect is disregarded (as is "the other side" of the story: any employer-inflicted coercion of labor workers).

Understood as an argument against the use of violence or physical force, whether or not sanctioned by state laws, the book provides a good overview of the inefficiencies arising due to a number of restrictions imposed on the labor market. In the case of labor unionism Block rightfully goes after the artificial increase in wages brought about through unions' [legal] threats of violence against employers – and the inevitable negative effects thereof. True to the economic analysis, Block argues that "unions cannot raise real wages, only distort them" and that "[u]nions are [...] notorious for undermining management's ability to do its job, which is to increase efficiency" (2008:100).

He also points to the fact that labor unions do not only act in the interest of labor workers against employers, but that there are strong incentives for union leaders, due to their privileged position, to "not only want higher wages for their members" but also "to squeeze every resource of the employer in order to make their union more attractive to prospective dues payers" (2008:103). The coercive labor union therefore distorts the labor market more than a general raise in wage rates would (without an equivalent increase in productivity), through adding incentives that ultimately will force employers to bear costs of union benefits and "marketing."

Even if it were the case that labor unions would act primarily as representatives of labor workers in conflicts with employers, it is argued that "[t]his is a very inefficient and costly way to settle problems which should never exist in the first place" (ibid). These conflicts would not exist in the free market, Block argues, since they arise due to labor unions pushing wages "above the level that a competitive free market would have brought through supply and demand" (2008:99), to which management "must respond by cutting back on production in order to minimize costs" (2008:103).

Block frequently falls back to almost a market equilibrium-based argument, using the free market as benchmark, against the coercive interference in the labor market – labor unions cause distortions through forcing employers to pay higher wages and undermining firm and production management's efforts to increase productivity and efficiency. It is an economic truth that "[w]ages and working conditions aren't set by firms"

(2008:111) but is determined solely, at least in the long run, by the productivity of each individual person in the market. Block's argument is no doubt firmly based in sound economics; as all economists know, in the competitive free market real wages depend only upon the productivity of labor, which means that the only way of increasing real wages is to increase productivity.

The problem is here that Block seems to partly forget the perspective he claims to have adopted in the analysis: libertarianism. Economists regularly analyze the effects of changes through holding all relevant variables but the one being studied constant; they normally use simplified models in which a single variable can be compared and contrasted with the benchmark equilibrium. Libertarians, on the other hand, guided by a libertarian *ethics*, would not find a strictly economic analysis satisfactory since it is too limited in scope and therefore would easily fail to notice important but "hidden" aspects and indirect causes of the problem; from a libertarian point of view the existing labor market is so far from being a free market that it is simply impossible to surmise that the distortions are the result only or for the most part of labor unions and union-sponsored, union-supportive regulations.

Murray Rothbard, a leading economist and political theorist in the libertarian tradition as well as in Austrian economics (and frequently cited by Block), has stated that a true libertarian is guided by "a passion for justice" and that such a passion requires "a set of ethical principles of justice and injustice which cannot be provided by utilitarian economics" (Rothbard, 1966:6). With Rothbard, therefore, we must conclude that the Blockian analysis once again falls short of being a manifestly libertarian analysis – it is primarily an economic analysis.

Even though a libertarian analysis would have no problem incorporating Block's conclusions, libertarians guided by a passion for justice would claim there is a much deeper and systemic problem than simply the existence of labor unions and the violence they make use of in the labor market. From a libertarian point of view there is as much of an "injustice" problem on the "other side" of the conflict: employers are not solely victims of unionized, state-sanctioned violence – they are also beneficiaries of a multitude of regulations. They may not normally use direct violent action against hired workers, but they are certainly not perfectly without blame. Corporations and employers enjoy state-sanctioned privileges in the market place just like labor unions. It is simply not the case that unions are villains and employers are not – they are both crooks, but in different ways and perhaps of differing degrees.

The problem from an economic point of view is here that corporations seem much more regulated at first glance, since labor unions enjoy obvious legal privileges and "labor protection" is explicitly and frequently advocated in political discourse. The common rhetoric used by political decision-makers and interest groups is almost without exception to the benefit of the worker – against "powerful" corporations in a hopeless David and Goliath kind of situation. It is therefore easy to assume that regulations are introduced as an attempt to politically strengthen laborers to balance the perceived "market power" of employers.

But regulations are as frequently to the benefit of employers. Regulations raise barriers of entry that protect existing actors in markets; taxes force increases to the working population (which pushes wages down) through making it impossible to afford choosing not to work; government investments in infrastructure and technology act as indirect subsidies to corporations; and the political system provides opportunities for corporations to "buy" their own laws from politicians eager to enrich themselves or gain support for reelection. These are all examples of interventions with direct effect in the labor market, but they are not as easily recognized as union violence. From a radically libertarian perspective they must be deemed at least as important and destructive as the effect of labor unions and union laws; libertarianism does not discriminate between different forms of injustice – all initiation of violence is equally illegitimate and immoral (Rothbard, 1982).

A radically libertarian view could identify a long list of interventions in the labor market that makes it fundamentally *un*free and the points on the list would be to the benefit *and* detriment to literally every actor in the market. Labor unions are to blame for the harm they do, but it is hardly the case that the market would function as a free market were only labor unions and union laws dropped from the equation. The libertarian conclusion would be that even if all the interventions analyzed by Block were removed, the market would still not to a large degree resemble a free market. The violations by or on behalf of firms and employers are absent from Block's analysis of the labor market – it seems to be guided by a one-eyed passion for justice.

To reinstate the free market and its institutions, libertarians would argue that all initiation of violence need be eliminated – systemic, formal, institutionalized and informal alike. And, as Block surely knows, for any market to be truly free it is necessary to abolish government.

3. *Summarizing Assessment*

Labor Economics from a Free Market Perspective is a provocative book: it is too libertarian to be an economics treatise while too firmly based in economic theory to be a libertarian exposition. As a libertarian anthology it is too limited in scope and "passion" to be a comprehensive investigation of the effects of aggressive violence in the labor markets, and as an economics work it is too polemical and provocative and "ideological" to be taken seriously by mainstream economists. So what is its place in the literature on labor economics?

It is hard to say exactly how to label this book, but it certainly fills a void in the intersection between economic and libertarian theory. In a sense, it proves that economic theory need not be as rigidly positive and lifelessly *wertfrei* as economists tend to believe – it is possible, and perhaps favorable, for economists to have a strong value-based motivation while carrying out economic scientific studies. Fundamental motivations for research are always and necessarily value-based, which means full disclosure of the scientist's value base would only provide explicit context for understanding, analyzing and criticizing the research – and the reasons for it. By being explicitly libertarian, Block does the reader a favor that economists generally seem determined not to.

This point is even stronger considering chapter 17, where Block recites a statement in support of the minimum wage signed by more than 650 prominent economists including Nobel Prize laureates. Unambiguously, economics shows that minimum wage laws only lead to unemployment and worsening of working conditions; in Block's words, "[e]very Basic Economics 101 textbook […] make this basic elementary point" (2008:160). The *reason* for the signatories to support this petition despite the obvious economic truths must therefore be normative – if these economists would have followed Block's example and disclosed their personal value based perspective, their signing of the petition would have been less befuddling. (Of course, the political intent would then be all too obvious.)

Judging from the signatories of the petition, the basic economic truths that Block recapitulates using his characteristically fearless approach and outspoken mode of expression obviously need repeating. And doing so with a distinctly libertarian flavor through explicitly focusing on market interventions *as violence* is refreshing and thought provoking. Even though it seems unintended, Block manages to prove that economic theory is generally compatible with libertarian political and moral theory without compromising

with the economic argument – and he shows that economics can indeed provide a strong argument for libertarianism.

References

Block, Walter. 1976. *Defending the Undefendable*. New York: Fleet Press.
Block, Walter. 1994. "Libertarianism and Libertinism." *Journal of Libertarian Studies* vol. 11, no. 1 (Fall 1994): 117-126.
Block, Walter. 1999. "Austrian Theorizing: Recalling the Foundations." *The Quarterly Journal of Austrian Economics* vol. 2, no. 4 (Winter 1999): 21–39.
Block, Walter. 2008. *Labor Economics from a Free Market Perspective: Employing the Unemployable*. Singapore: World Scientific Publishing Company.
Hoppe, Hans-Hermann. 1993. *The Economics and Ethics of Private Property: Studies in Political Economy and Philosophy*. Boston: Kluwer.
Hume, David. 1739-1740. *A Treatise of Human Nature: Being an Attempt to introduce the experimental Method of Reasoning into Moral Subjects*.
Mises, Ludwig von. 1949. *Human Action: A Treatise on Economics*. New Haven: Yale University.
Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
Pareto, Vilfredo. 1971. *Manual of Political Economy*. Translated by Ann S. Schwier. Edited by Ann S. Schwier and Alfred N. Page. New York: A. M. Kelley.
Rothbard, Murray N. 1962. *Man, Economy, and State: A Treatise on Economic Principles*. Princeton, N.J.: Van Nostrand.
Rothbard, Murray N. 1966. "Why Be Libertarian?" *Left & Right: A Journal of Libertarian Thought* vol. 2, no. 3: 5-10.
Rothbard, Murray N. 1982. *The Ethics of Liberty*. Atlantic Highlands: Humanities Press.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 248-252

Commentaries on Gordon and on Bylund

Walter Block
College of Business Administration
Loyola University New Orleans
wblock@loyno.edu

1. *On Gordon*

Consider most people whom you have been friends for a long period of time. Typically, and this is certainly true in my case, one is hard put to recall exactly how you first met them. If they are good enough friends, as in the case of David Gordon and me, recollection of this first meeting vanishes into the mists of time. One simply cannot remember.

David is different in this regard, as he is in so many others. I *distinctly* remember my first meeting with him, (almost) exactly what he said to me, and my own reaction, which was one of profound astonishment.

He said something to me very much like this: "You're Walter Block. It's great to meet you. I have been for a long time a fan of your writings. For example, in your publication with title A, published in the year B in journal C, you said on page D, the following. Whereupon he would offer me a long quote from my publication. Then he would say, "in your publication with title E, published in the year F in journal G, you said on page H, the following".

Whereupon he would offer me another long quote from another of my publications. He continued this pattern for quite a while, offering such renditions of some half dozen examples of my publications. His quotes of me all sounded plausible; for all I knew, they were verbatim. I had no idea of whether he was correct in any of this. I was too shocked to be able to make any such determination. Indeed, my recollection is that my jaw hung low in amazement.

Why am I relating this story from my past? It is because I was brought in mind of it by my present experience of reading what David said about my book *Labor Economics From a Free Market Perspective*. With the experience of several decades, now, of David's friendship and companionship, I have no doubt that with his photographic memory, the cites and quotes he reeled off to me at that long ago first meeting were entirely accurate.

I have a similar reaction to his review. I have no doubt that he has memorized the entire book. But more. It is my understanding that there are a few people on this earth who are capable of such feats. But David combines this with an under-

standing that is rare indeed. He has plumbed the depths of this book, in a way that I myself, the author, feel inadequate to do.

He says he has concentrated on the "central theme" of my book. Not so, not so. He has done far more than that. He has *pierced* the central theme of my book. If David were a bullet, and my book a piece of heavy reinforced steel, there would now be a hole in this publication of mine big enough to drive a truck through.
Put it this way. If I were a murderer, and David a detective, I would not at all appreciate him being assigned to the case. All I can see is that in this intellectual battle of ideas we Austro libertarians find ourselves in the midst of, I appreciate every day of it knowing that David is on my side of it.

2. *On Bylund*

I am very grateful to Per Bylund for his commentary on my book. He does me great honor with his thoughtful critique of it. I am a great believer in the intellectual benefits of criticism, of which his essay is a splendid example. Let me reply to his under two headings: non substantive and substantive.

Several times in his review Bylund makes assertions without offering examples. For example, he refers to my "typically Blockian manner, very outspoken, shocking, and – sometimes – even infuriating." Just out curiosity, I would like to know to what, specifically, he is referring. He mentions my "provocative language and somewhat unorthodox examples" (thanks to the modern miracle of word search, I am able to report that Bylund employs the word "provocative" no fewer than five times), again without satisfying my (or anyone else's, I suppose, inquisitiveness). This author maintains that my "work […] is too polemical and provocative and 'ideological' to be taken seriously by mainstream economists." Well, been there, done that; I am very accustomed to having my efforts dismissed by neoclassicals, but what, pray tell, did I do *this* time so as to call forth this reaction? I am left in the dark.

Also, under this heading I cannot resist rejecting his several criticisms of the books omissions, not commissions. For example, (note, I am very specific in my criticisms of Bylund), he says, "but (Block's) […] depth of the analysis of such intervention in the labor market is insufficient – the analysis is too limited and does not take into account all major effects of market regulation." Well, of course. No book can be "unlimited." It seems a bit harsh to condemn a book for not taking "take into account all major effects of market regulation," when that was not the avowed goal of the author. As well, my commentator states: "And, as Block surely knows, for any market to be truly free it is necessary to abolish government." He does so in the context where he seems to be blaming me for not men-

tioning this elemental and basic truth in *Labor Economics from a Free Market Perspective: Employing the Unemployable*. Again, not every book can say every correct thing.

Bylund continues in this vein: "From a libertarian point of view there is as much of an 'injustice' problem on the 'other side' of the conflict: employers are not solely victims of unionized, state-sanctioned violence – they are also beneficiaries of a multitude of regulations." And again: "The violations by or on behalf of firms and employers are absent from Block's analysis of the labor market – it seems to be guided by a one-eyed passion for justice." Here again I am being blamed for what I did *not* say, not for what I *did* say. Remember, this is a book on *labor markets*. Were I to have written one on the corporate sector, I certainly would have made the points so eloquently put forth by Bylund. And, as to the relations between employer and employee in the *labor market*, it is my view that violence perpetrated by companies (Pinkertons' etc.) was overwhelmingly a reaction against prior violence perpetrated by organized labor. If unions were to limit them selves to mass quits, and not set up set ins or pickets, which violate the private property rights of the company, there would simply be no *need* for them to resort to violence.

One might as well blame the bible for not mentioning physics. One might as well hold mathematicians guilty for ignoring Shakespeare.

Now for substantive matters. Mr. Bylund spends what I consider to be an inordinate amount of effort on the normative-positive distinction, the fact value dichotomy, the difference between economics, on the one hand, and political philosophy, particularly libertarianism, on the other. Which animates this book? In my opinion, *both* do, only in different parts of the book. Remember, this volume is a compilation of many previously published articles, which now comprise its chapters. Well, on *some* of these occasions I am involved in the one perspective, on others of these occasions the other. Even if this were not the case, there is nothing wrong or improper with speaking from these two perspectives within the same essay, even a short one, provided, only, that the Humean distinction is not violated. Try as I may, I do not see from Bylund's commentary any examples of any such violation: that is, me deducing an "ought" from an "is," for example, saying something like "the minimum wage creates unemployment for the unskilled, therefore it is wrong." If there is anyone who commits a sin against the normative positive distinction, it may well be Bylund, who says: "economics as a scientific discipline is founded on a utilitarian philosophy of what is universally good." In my view, economics is founded on no such thing. For the Austrians, at least, the dismal science, is, instead, predicated upon human action. In this regard, the following statement can only be considered problematic: "economic theory is generally compatible with libertarian political and moral theory." To the contrary,

economic theory is in the positive realm; libertarianism, in contrast, lies entirely within the arena of the normative. Never the twain shall even meet, let alone with "compatible" with one another, at least according to Bylund's authority on this subject, Hume.

In addition, here are several points of divergence between me and my critic. I quote from him, and comment:

"Economics in general, and especially the Austrian tradition, is deductive in nature." But, this is not exactly true. Bylund is entirely accurate with regard to praxeological or Austrian economics, but as for "economics in general," by which I presume he means mainstream or neo classical economics, this is an inductive, not a deductive, "science."

"Austrian economics rejects statistical methods and empirical studies as means to learn about economic truths and hence adopts a purely deductive approach." Yes, Austrian economics rejects statistics and empirics as a *test* of praxeological law. But these methods are not at all eschewed when it comes to *illustrating* basis economic axioms. This may sound like hair splitting, since it is a popular fallacy about Austrian economists that we rebuff *any* connection with statistics or empirical studies. Not so, not so.

"Any ideology is by definition normative and therefore so is libertarianism." I can't see my clear to agreeing with this. The suffix "ology" means, merely, "study of." For example, biology is the study of life; archeology is the study of (pre) history of human culture; geology is the study of ancient rock formations. None of these are normative, even though they are all ideologies, e.g., the study of ideas.

"libertarianism […] only limits individuals' actions to anything that does not do direct harm to other individuals." Close, but no cigar, here. Boxers do "direct harm" to each other, by punching each other in the mouth. A asks B to marry him. B refuses. A commits suicide as a direct result. This would appear to constitute "direct harm." Yet, boxing, refusing marriage proposals, are certainly compatible with libertarianism. Libertarianism doesn't prohibit "harm," only rights violations, which none of these are.

"Austrian economics studies and predicts the negative effects of interventions in the market place." Nor really. Rather, any prediction that Austrian economics makes are contrary to fact conditionals: *If* all else remains the same, then X will lead to Y. But we are never in a position to assert that all else stays constant. (For more on this see Hulsmann, Jorg Guido. 2003. "Facts and Counterfactuals in Economic Law." *The Journal of Libertarian Studies*. Vol. 17, Num. 1, pp. 57-102; http://www.mises.org/journals/jls/17_1/17_1_3.pdf)

"[…] libertarianism, based on the non aggression principle, cannot espouse any other economic 'system' than a free market economy – all alternative ways of economic organizing would necessarily violate the fundamental principle. The free

market and libertarianism are therefore, in substance, two sides of a coin." Not quite. Voluntary socialism is also compatible with libertarianism. Organizations such as the monastery, nunnery, kibbutz, commune all adhere to the doctrine of "from each according to his ability, to each according to his need." Even the typical family engages in this practice. The three year old girl eats in accordance with her need, not her ability to earn income.

"government investments in infrastructure and technology act as indirect subsidies to corporations." Not so fast. Compared to what? Take roads, as an example. Had the government not built them, undoubtedly, private enterprise would have done so. (Block, Walter. 1979. "Free Market Transportation: Denationalizing the Roads," *Journal of Libertarian Studies: An Interdisciplinary Review*, Vol. III, No. 2, summer, pp. 209-238; http://www.mises.org/journals/jls/3_2/3_2_7.pdf). Entrepreneurs in the private sector would have done a better job, at a lower price. Thus, compared to the operation of free enterprise system, these government investments act as detriments, not subsidies.

Despite these divergences of views, I am very grateful to Bylund for honoring my book with his otherwise incisive comments.

VARIA

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 256-273

Moral Facts, Possible Moral Worlds and Naturalized Ethics

Fasiku Gbenga
Department of Philosophy
Obafemi Awolowo University, Ile-Ife.
platoife@oauife.edu.ng

ABSTRACT
Given his commitment to the project of naturalizing every normative aspect of philosophy; reducing its a priori content to some sort of empirical enterprise, Quine's inroad into moral philosophy is expected to set the stage for the project of naturalizing ethics. However, Quine argues that ethics is methodologically infirmed. Hence, the hope of naturalizing ethics hits the rock. This paper aims at advancing the project of naturalizing ethics by an attempt to settle, *in a way different* from the postulations of Flanagan and White, foremost commentators on Quinean ethics, Quine's charge of methodological infirmity.

1. *Introduction*

Since 1978, when Quine published his only paper on ethics entitled "On the Nature of Moral Values"[1], quite unlike many of his publications in other areas of philosophy, the level of debate generated by this essay is quite low[2]. Possible reasons that could be adduced for this are, first, Quine is delving into a strange land and had probably not said anything controversial enough that is worthy of academic dispute. Second, Quine's aim is to show that, unlike other areas of discourse, given the specialty of Ethics, its method makes it to be outside the 'naturalized world' and having shown this, there is nothing more to debate. While my first postulation is trivial, hence, indefensible, the second postulation is cogent[3]. Scholars who had written on Quinean Ethics so far are sharply

[1] The paper was first published in Goldman A.I. and Kin J. (eds.) (1978: 37-46). The version referred to in this paper is in Quine W.V. (1981:55-66).

[2] The only known substantive articles on Quinean Ethics are four. These are Flanagan O. J. (1982: 56-74), White M. (1986: 649-662), Gibson Roger F., "Flanagan on Quinean Ethics" (the version of this paper I used in writing this paper is unpublished.) and Quine's "Reply to White". (1986: 663-665).

[3] Several arguments are being offered to underscore the appeal of naturalism. The possible truth of these arguments exacerbates the need to incorporate ethics into the naturalist

divided on whether or not Quine is right in arguing that on the basis of methodology, ethics and science are quite different in all ramifications. Flanagan and White hold that for Quine to be a consistent naturalist, his conclusion that ethics is methodologically infirm is unwarranted, hence, he ought to continue with the project of naturalizing ethics. Gibson, on the other hand, supports Quine in arguing that on the basis of methodology, ethics and science do not belong to the same boat.

In what follows, I shall attempt to advance Quine's project of naturalizing ethics by an attempt to settle, in a way, different from the postulations of Flanagan and White, Quine's charge of methodological infirmity against ethics. In what follows, a short explication of Quine's account on the genealogy of moral values is carried out to establish the point that moral value is generated in the same way as it is done in science and other naturalized discourses. This is followed by a concise exposition of Quine's argument for the charge of methodological infirmity against ethics. In an attempt to advance Quine's project of naturalizing ethics, the difference between the notions of possible worlds and actual or natural world would be used to explain the existence of moral facts to which moral judgments would correspond. This is used to show, in the conclusion, that ethics as compared to science is not methodologically infirm.

2. *The Technology of Moral Values*

Ethics is a branch of philosophy that is concerned with the body of principles or standards of human conduct that govern the behavior of individuals and groups. It is considered a normative science because it is concerned with the norms of human conduct, as distinguished from formal sciences such as mathematics and logic, physical sciences such as chemistry and physics, and empirical sciences such as economics and psychology. Ethics arise not simply from man's creation but from human nature itself making it a natural body of

programme. See Virginia Held, (2002: 7-24), for a few of the arguments on the appeal of naturalism. Quite pointedly, in this article, Held, p. 8, also argued that "naturalism holds out hope of philosophical knowledge that will progress along with advances in specialized scientific inquiries, and of moral knowledge that will advance along with scientific knowledge".

laws from which man's laws follow. As Schueler observes, "The human conceptual apparatus, including that part of it involved in making and acting on moral judgments, is somehow instantiated in the brain and nervous system"[4]. Ethics is a natural, scientific and technical phenomenon. This suggests that ethics exists as parts of the natural structure of the world. It evolves on its own course, as response to the other structures of the world. It is parts of the supporting pillars of the natural world, without which the world would have been different. Put differently, ethics is a metaphoric walking stick that human beings, one of the structures of the world, need to stand, withstand and walk through the other features of the world[5]. Hence, as speculated in the philosophy of Democritus, "struggling to survive against hostile forces in his environment, man is compelled to associate himself with other men; hence speech. He is also compelled to *learn* from *experience*; hence the *mechanical arts*."[6] This compulsion is explained by the fact that ethics, as one of the fabrics of the world, complements other features of the world. It is, therefore, a natural phenomenon subscribed to by every rational human being. Hence, just as every other structures of the world is studied Ethics, the moral institution qualifies as a natural edifice, which as Quine notes, is "to be studied in the same empirical spirit that animates natural science".[7]

One important characteristic that distinguishes human beings from other species of animals is the ability to make rational and informed choices. These choices are motivated by values. Hence, Quine explicates the relationship between the capacity to make rational and informed choices and the value of the choices made. For Quine, this capacity and the value made are intertwined. Encompassed in the concept of capacity is what Quine call 'belief'. For Quine, 'belief and valuation intertwined'.[8] In the belief aspect are the epistemological components of the ability to make rational and informed choices. The epistemological components among other, "involves standards of perceptual

[4] G.F. Schueler, (1996: 315).

[5] The same point of the naturalness of morality was emphasized by Annette Baier. For details see, Annette Baier, (1996: 5-17)

[6] Gregory Vlastos (1946: 54). (I deliberately put these concepts in italics. These concepts would be used to explicate the point that ethics is a natural, scientific and technical phenomenon.)

[7] Quine, W.V.O. (1969: 26)

[8] Quine. W.V.O, (1981:55)

similarity: some of these standards are innate, others are acquired."[9] Other components include awareness of the object of value by the subject. The choices that are made are consequent upon the epistemological component of belief. Moreover, according to Quine, our value involves pleasures and pains. In this respects too, there are innate likes and dislikes as well as acquired likes and dislikes, which guide our choices. Since rational human beings would naturally want to maximize pleasure and avoid pain, the standard of perceptual similarities becomes an essential instrument in evaluating episodes appropriately, either as pleasurable or painful. Hence, "the drive to increase or decrease the similarity will…vary with the degree of pleasantness or unpleasantness of the earlier episode."[10]

According to Quine, "the similarity standards are the epistemic component of habit formation, in its primordial form, and the reward-penalty (pleasure-pain) axis is the valuative component."[11] The similarity standard becomes the instrument that shapes human's thoughts and world-views. This is clearer as Democritus notes, "the nature of the soul is not fixed by original pattern of the soul-atoms. This pattern itself can be changed: Teaching re-forms a man, and by re-forming, makes his nature." [12] This explains the Democritus's dictum "teaching that makes nature". What could be derived from this is the point that epistemology is prior to metaphysics. This is because it is what you know that shapes your world. However, for Quine, the relationship between epistemology and ethics is complementary. For him, "(c)learly, all learning, all acquisition of dispositions to discriminatory behaviour, requires in the subject this bipartite equipment: it requires a similarity space (epistemological component) and it requires some ordering of episodes along the valuation axis(ethics), however crude."[13] It is this exercise of fulfilling these bipartite requirements that exacerbates the science and technicality of moral discourse. The similarity space and the ordering of episodes are being studied, progressively changed and elaborated through scientific method of induction, and eventually, hypothetico-deductive method.

---

[9] Gibson. R.F. "Quine on Ethics" unpublished p.5
[10] Quine. (1974: 28)
[11] Quine. (1981: 55)
[12] Gregory Vlastos. (1946: 54)
[13] Quine. (1981: 56)

An example will suffice in illustrating how these methods of natural sciences are used in realizing values. If we discover that a particular kind of act or thing, say X, always produce a desired end, say Y, and since, every normal human being would always want a repeat of what is desirable, it is probable that there would be desire to repeat X in order to get Y. Hence, for example, it is on the basis of inductive reasoning that I infer that a new computer system will serve me well on the ground that I got very good service from a number of computer systems earlier purchased from the same manufacturer. Again, if a new book by a certain author is introduced to me, I infer that I will enjoy reading it on the basis of having read and enjoyed other books by that same author. It is, also, by induction that I reason that my car, made of metal would have a dent if it hits a harder object, this is because I have observed several cars made with metals that got dents when hit against harder objects. Having experienced an event or a thing being followed by the same effect always, the scenario is now believed to be part of nature. The experiences are summarized into general laws or universal generalizations. For example, having observed that cars made with metal get dents when hit against harder objects, I then generalize that 'All objects made with metal when hit against a harder object will have a dent'. On this basis, once I see a car, made of metal, hit against a harder object, I deduce, without further observation that the car must have a dent. This is clearly an example of the hypothetico-deductive method.

In the same vein, this scenario also obtains in terms of valuations.[14] According to Quine, when "we learn by induction that one sort of event tend to lead to another that we prize" (It is important to note as earlier remarked, that this inductive process, is an epistemological component that human beings possess innately or acquired) and then by a process of transfer we may come to prize the former not only as a means but for itself"[15]. This means that on the issue of values, reasoning starts from induction, by observing instances of what event or thing is valuable and or otherwise. On the basis of these observations, these events or things are valued for themselves, not because they lead to other

[14] It is important to note that valuation is an inevitable exercise for beings. As Shirk notes, 'value is assigned to people, places, acts, sensations, or thoughts. This is an exercise that is almost inevitable in human affairs. For details on value and different dimensions on value, see Shirk Evelyn (1965)

[15] Quine (1981: 57)

values. These events or things, valued for their own sake, therefore, become the conditions of assessing other kinds of event or thing.

This, according to Quine, also obtains in ethics. As he reasons, "many sorts of good behaviour have a low initial rating on the valuation scale, and are indulged in at first only for their inductive links to higher ends."[16] These good behaviours are arrived at through inductive reasoning, and they are used to generate higher ends, which are valued for themselves. The good behaviours, therefore, become means and the higher ends become the end. Hence, the good behaviour forms the premises of an inductive argument in which the higher end is the conclusion. The more the instances of the good behaviour are obtained, the more the higher end is confirmed. The good behaviour becomes a moral value if it is turned into an end-in-itself or a higher-end; which is demanded for its own sake, not as means to an end. This is done by making the good behaviour a general statement or a universal generalisation in a hypothetico-deductive method of reasoning. For Quine, it is this process of "transmutation of means into ends… (that) underlies moral training"[17]. Take for example, in Yoruba culture, if I prostrate to greet someone, my action will be applauded a good behaviour. This good behaviour becomes a premise of a higher end, say, respect. For Quine, the act becomes a moral value when the good behaviour, which is a mean to an end, transmuted into an end, and is therefore valued for itself and no longer as a means to an end. So, the act is performed habitually without experiencing the slightest applause. Hence, a general law ensued, through which other similar behaviour is assessed. Thus, consider this example:

(1) "it is a good behaviour for Yoruba male child to prostrate while greeting an elderly person"

(2) Biodun is a Yoruba male child

(3) Biodun prostrated while greeting his father

(4) Therefore, Biodun's act is a good behaviour.

The above is an instance of a hypothetico deductive method of reasoning. (1) is a hypothetical statement under which (2) and (3), the initial conditions or

[16] Quine (1981:57)
[17] Quine (1981: 57)

instances, are subsumed, and once these two hold, (4) the conclusion is derived. It is through this system of reasoning that moral value is produced among human beings. This, indeed, is a technical affair, hence, as Quine remarks, "good behaviour, insofar, is technology"[18]. Quine's distinction between moral value from other kinds of values is summarized by Gibson: "moral values, as opposed to moral values, are 'irreducibly social', i.e., they are oriented towards the satisfactions of others"[19] The important point that is being underscored in this section is that moral values and moral standards are derived in the same way as scientific theories are derived.

3. *Ethics and the Charge of Methodological Infirmity*

The process of transmutation of means to ends as explicated above suggests that ethics follows the same pattern of reasoning in establishing moral values, as is the case in natural sciences. The establishment of this point should ordinarily provide a ground for accepting ethics as belonging to the naturalist family. However, Quine argues that the parity between ethics and science does not hold in respect of the method required in settling disagreements. For him, when disagreements occur on moral matters, "one regrets the methodological infirmity of ethics as compared with science".[20] He argues that there are empirical events or states of affair, which serve as empirical footholds of scientific theories. For example, there is the actual event of water getting boiled at 100°c, which confirms or corroborates the scientific principle that "water boils at 100°c". Similarly, there is the actual event of deliberate killing of innocent persons, which serve as empirical foothold of the moral code, 'murder, i.e., deliberate killing of person, is bad'. The problem is that "whereas, (in science) we can test a prediction against the independent course of observable nature, we can *judge the morality* of an act *only* by our moral standards themselves."[21] So, in case of disagreements about whether or not water boils at 100°c, we can point to the physical fact of the actual event of water getting boiled at 100°c, as the evidence for the justification of the prediction embedded

[18] Quine (1981: 57)
[19] Gibson (Unpublished:10)
[20] Quine (1981:63)
[21] Quine (1981: 63). Italics mine.

in the scientific theory – water boils at 100°c –, regrettably, there is no such fact of badness or wrongness, out there in the world, that would serve as evidence in the judgement of the act of deliberate killing of innocent persons as being morally bad or morally wrong, other than making a recourse back to the moral standard – 'deliberate killing of person is bad'. Hence, "science, thanks to its link with observation, retains some title to a correspondence theory of truth, but a coherence theory is evidently the lot of ethics[22]. This is because there is no observable entity that ethics can be linked to in the world.

As Quine further notes, "extrapolation in science, however, is under the welcome restraint of stubborn fact: failure of prediction. Extrapolation in morals has only our unsettled moral values themselves to answer to, and it is these that the extrapolation was meant to settle"[23]. It is on the basis of the unavailability of observable entity in ethical discourse that renders ethics incompetent as being a natural enterprise. Ethics belongs to the normative discourse, which is bound by its internal strings of, mostly debatable and non-objective, laws and theories, which are often the sources of moral disagreements and moral conflicts.

To repeat, Quine's charge is that in an attempt to resolve moral disagreements or moral conflicts, there are no 'empirical checkpoints', which are the solace of the scientist. Hence, ethics and science differ in a major respect, thus, the two belong into different boats. In what follows, we shall examine arguments for and against this position. In the end, an attempt is made to make a case for the ontology of moral facts, which when observed, would break the circle of reference to moral standard in order to justify moral judgments, and, hence, bridge the gap between science and morality.

There are at most, two known naturalists who had commented on the subject of Quinean ethnics. They are Flanagan[24] and Morton White[25]. For its sharp relevance to the dimension of arguments being sketched, I shall be concerned

---

[22] Quine (1981: 63)

[23] Quine (1981: 65)

[24] Flanagan O. J. (1982: 56-74)

[25] Morton White (1986: 649-662). White's suggestion is that feeling would play the same role that physical facts play in observation. So, ethical judgment would then correspond to feelings in order to be justified. This suggestion was, however, refuted by Quine. For him, our feeling is part of the moral evaluation that needs justification. It is what conforms to this moral evaluation in the natural world that renders ethics methodologically infirmed. For details, see Quine, W.V. (1986: 663-665)

with Flanagan's attempt to advance the project of naturalizing. My understanding of Gibson's critique of Flanagan raises some issues, which needs further scrutiny. Notwithstanding these issues, Gibson's critique of Flanagan obviates the charge of methodological infirmity leveled against ethics.

Flanagan's contention of Quine's charge of methodological infirmity in ethics is not aimed at removing the infirmity. Rather, Flanagan argues that the charge is unwarranted because such a problem is not peculiar to ethics, but is a characteristic of all significant discourses. In the main, Flanagan argues that following Quine's holism, it is no longer fashionable for science to rely on observation as the paradigm of objectivity. Just like what obtained in ethics, the coherence theory of truth is also the lot of science, hence, it makes no sense to distinguish between science and ethics on the basis of methodology.

Gibson's challenge of Flanagan's position is that the latter is based on a misconstrued notion of the nature and scope of Quine's holism. Gibson shows that while Flanagan's conception of Quine's holism as, in summary, 'that all checks are ultimately intersystemic', is too broad, and therefore, erroneously concluded that ethics and science belong to the same boat, the correct conception of Quine's holism is a form of mitigated holism, which allows some sense of distinction between observation sentences and other kinds of sentences. The former have their own meaning derived from observation, while the latter derive their meanings from being members of a system. The crux of Gibson's challenge to Flanagan's understanding of Quine's holism is the failure to make this distinction. With this distinction, the gulf between science and ethics remains. What Gibson did not show, however, is that the observation sentences are mainly the lot of science. In other words, Gibson ought to show that ethical judgments cannot behave like observation sentences. In response, Gibson's acceptance, following Falanagan's that ethical statements can also have relation to experience, could suggest that ethics, just like science, has some title to the correspondence theory. However, Gibson thwarted this line of thought by reiterating Quine's earlier charge, though in another language, to show that Flanagan's suggestion, that consequence of a moral practice would be the observable fact that such ethical statements would correspondence to, would not suffice. Gibson insisted that in relation to moral values, there has to be objective facts, indisputable facts, which would exhibit the morality of the act: the wrongness or goodness, as the case may be, of the act, which ethical statements would correspond to. Without these facts, Quine's thesis of methodological infirmity remains.

It is pertinent to note that the problem explicated above is the root of the dispute between the proponents of the moral realist and moral anti-realist. The problem is described as follow: "Physicists appeal to the presence of protons to explain the observation of vapor trails in a cloud chamber. The vapor trails act on our visual system to produce an observation of them, and, with background knowledge, an observation that protons have passed through the cloud chamber, producing the vapor trail. But in the moral case there is nothing present in the objective facts to act on perceptual systems to produce the observations about moral rightness. Subject-side factors alone suffice to account for whatever moral observations or beliefs are generated in the situation"[26]. As a response to this anti-realist position, the moral realist argues to establish the ontology of moral facts. However, I believe that the threshold of the argument for the unity of ethics and science on the basis of the method of settling disputes is by establishing the ontology of independent, objective and moral facts which exists as parts of the fabrics of the world. Once established, it is to these facts that moral judgements or statements would correspond. In what follows, I shall attempt to articulate arguments that establish the ontology of moral facts.

4. *Moral Facts as the Threshold of Naturalized Ethics*

"Can't you *see* that *this* is *wrong*?" "Could you *imagine this* being *right*?" "How could you have done such *a thing* like *that* (which is *wrong*)?" These are questions that appeals to the *fact* of the *wrongness* of a particular act. In each case, the questioner invites the listener to *see*, *imagine*, and *consider* the fact that the act in question is wrong. What is being demanded is to, like natural fact which is out there, independent of the observer; establish the ontology of the fact of *wrongness* as an observable, objective entity that exists independently of the moral subjects.

What I propose, however, is that moral fact is an 'entity' that exists in all possible world, in which there is no world in which moral discourse operates and the fact would be denied. By this I mean that the 'entity' *wrongness* in a moral judgment, such as, 'This act is wrong', 'exists', not as entities in the 'actual world', and observable through empirical apparatus with which natural facts are observed, moral facts are kinds of entities that are 'observable' as a possible

[26] Rottschaefer, W.A. (1999: 1)

entity in every 'possible moral world'. In this possible moral world, these moral facts are 'observable' giving its stipulated laws and principles. It is this entity that is referred to when we say that 'an act, say x, is wrong. In this case, x is wrong if and only if x is wrong in every possible world in which x exists'. The point is that x being wrong in every possible world is the fact that is being appealed to in the moral judgment: 'x is morally wrong'. If there is a possible world in which x would be right, then the moral judgment that 'x is morally wrong' would not correspond to any moral fact. So, when I say that 'Can't you see that this act is wrong?' I am only inviting you to 'observe' the fact that there is no possible world in which the act exists and it is morally right. If my hearer could justifiably show that there is a world in which the act is right, then my moral judgment would not correspond to any fact.

What derives from this understanding of moral fact is that the actual wrongness of a morally wrong act is not an empirical entity; it is a fact because it is not corrigible in the present and any possible world. Though, the fact is not observable in the same sense in which natural facts are observed, they are observed by all the subjects concerned by searching through the entities in all possible worlds in order to see there is no fact that run contrary to the moral fact.

This account of moral facts above rests heavily on the notion of possible world. It is also based on a distinction between 'world actual' and 'morally possible world'. A detailed discussion on the notion and problems associated with possible world is beyond the scope of this paper. I shall, however, offer a brief discussion of these notions in order to explicate my position.

'Possible world' is one of the numerous terms used by philosophers to elucidate, analyse and proffer solutions to a number of philosophical problems.[27] However, the question 'what is a possible world?' is a philosophical problem that has no consensus solution. However, there are, among others, two prominent positions. The first is the extreme realist position, largely attributed

[27] The notion of a possible world is not new in Philosophy. The Pre-Socratics had in one way or the other postulated the idea of possible worlds in their speculations about reality. Of particular interest is Parmenides idea of two ways of the world and the Atomists': Leucippus's and Democritus's idea of 'unboundedly many worlds'. For detailed account of the Pre-Socratic conceptions of possible worlds and what they use it to achieve, see Kirk, G.S. et al, (1983). In the contemporary epochs, the notion is commonly used by philosophers in modal logic to elucidate the distinction between necessity and possibility. The notion is prominent among Kripke, Plantinga and David Lewis to mention just a few.

to David Lewis, which maintained that a possible world is another real or concrete world just like ours. On this view, the notion of possible worlds is not just a philosophical tool useful for the purpose of elucidating philosophical arguments or claims. Possible worlds are real in some way. In this conception of possible world, what makes worlds distinct is that they are spatio-temporally separated from one another. In other words, every way that a world could have been is a way that some existing physical world really is. So, possible worlds are real worlds and they actually exist in the same sense the real or concrete world we inhabit exists.[28]

The other position is the moderate realist position supported by Alvin Plantinga, A. Adams and others, who have claimed that a possible world, is nothing but an abstract entity, and does not really exist. For the moderate realist, the notion of a possible world is merely a useful philosophical tool for making arguments. The moderate realist position is that the notion of a possible world refers to the ways we imagine that the world could have been different from the way it is. A possible world is a way a universe might have been. Possible worlds are counterfactual states of affairs. States of affairs are abstract entities that such phrases as

(1) 'Socrates died after drinking poison'

and

(2) 'Socrates having lived after drinking poison'

refer to. Some states of affairs obtain, others do not. Proposition (1) refers to a state of affairs that obtained and proposition (2) refers to a state of affairs that does not obtain. Though the latter does not obtain, it is a possible state of affairs. It is different from a logically impossible and either causally or empirically impossible state of affairs.[29] The concept of a state of affairs is used to define what a possible world is. We imagine some states of affairs as being different from what they in fact are. These different states of affair are referred to as possible worlds.[30] As opposed to the extreme realist view that possible worlds are concrete worlds that exit just as our world exists, the moderate

---

[28] Lewis, David (1986:2)

[29] A state of affairs that is logically impossible if it does not respect the law of contradiction. For example, a state of affair such as 'it is raining and it is not raining' is logically impossible. A state of affair is causally or naturally impossible if it stipulates what cannot be physically achieved. For example, the state of affair 'Obasanjo having swum through all seas in the world' or 'Obasanjo having spent 1 million years on earth'.

[30] Alvin Plantinga (1978: 44).

realists assert that possible worlds are possible or imagined state of affair. It is how a world could possibly have been.

Following the extreme realist arguments, it may mean that there is no difference between the actual world and other possible worlds. This is because, for them, 'the actual world' means 'the world where I am located', and each possible world is actual from the point of view of its inhabitants. The term 'actual' is an indexical term like 'I'. It means 'part of the world of which I am a part' or 'part of the world of which this utterance is a part'. What Lewis means by the claim that 'actual' is an indexical is that actuality is not a necessary property of a particular world. According to Lewis, "surely, it is a contingent matter which world is actual. A contingent matter is one that varies from world to world. At one world, the contingent matter goes one way; at another, another. So, at one world, one world is actual; and at another, another. How can this be absolute actuality? – The relativity is manifest!"[31]

This means that every world is potentially actual; actuality is a property relative to all possible worlds. An Actual world is only one of other possible worlds. It is called an "actual world," not because it is different in kind from other possible worlds, but because it is the world in which the speaker inhabits. To the inhabitants of other worlds, their worlds are actual.[32] Put differently, for Lewis, the word 'actual' and the phrase 'the actual world' being indexicals are rigid designators. Lewis's argument is that when I utter the word 'I', it denotes me. Innumerable number of persons could utter the word 'I' at the same time; the referent of the word is each individual who utters the word.

However, Lewis' view about actuality rests on the realist assumption that there are other worlds that exist just as the world we live in and the inhabitants of these worlds are just as we are; it is this assumption that needs to be proved. The argument about indexicality of 'actual' and 'actual world' merely shows that all the possible worlds are potentially actual. There is a difference between a potentially actual world and a real world. The real world is different because apart from being actual, it is real, while the other actual worlds remain at the level of potentiality. However, Lewis' account would not admit this distinction. This is because, for him, there is no difference between worlds. All worlds are the same. The moderate realists would accept the distinction, and this makes their account more plausible. Let us explore moderate realism on the actuality of

[31] Lewis, David (1986: 94)
[32] Lewis, David (1973: 86)

possible worlds.

The moderate realists states that the actual world differs in ontological status from merely possible ones in that it is the only world that obtains. For Plantinga, 'an actual world is a maximal possible world that obtains.'[33] This implies that an actual world has the same status as the other possible worlds, but it is special because it obtains. A possible world that obtains is one that actually exists. While other possible worlds remain non-actual, the actual world is real. An actual world is a description of a state of affairs that is real, different from 'how things could have been'. 'How things could have been' is the description of possible worlds. So, possible worlds are different from actual worlds. The latter is real, while the former is merely possible. For moderate realism on possible world, only one world obtains, and it is that world that is named actual world. All other worlds that do not obtain exist as possible worlds.[34]

Given this understanding, the natural world is the actual world. It is the picture of how the world is actually is. The possible worlds are how the world could have been. A possible moral world is not an actual moral world, it is how a moral world could have been; it is, following the moderate realist position, an imagined or possible moral state of affair.

The point I am canvassing is that the moral facts are real in the sense that they exist in all possible moral worlds. A moral judgment is tested against a moral fact that exists in all possible worlds; it is a fact because it is found in possible worlds, and there is no world in which its contrary is found. If, however, there is a possible world where it is justifiably shown that the moral fact does not exist, then the moral judgment would not correspond to any moral fact. Such a judgment is therefore false. In this respect, moral fact, the wrongness in a moral judgment – deliberate killing is morally wrong – is a fact, if and only if, it is shown that there exists no possible moral world in which the act is of deliberate killing is morally right.

The mistake the naturalists like Quine makes is to treat moral fact as an entity, like neutron, proton, neurons etc, all of which are physical entities that exist in the physical or actual world. Moral facts are facts of a kind which exist in all possible worlds. I would have agreed with G.E. Moore[35] that those who are

[33] Alvin Plantinga (1978: 45)
[34] Alvin Plantinga. (1978: 47)
[35] Moore G.E. (1971: 16-17)

looking for moral facts among natural facts commits naturalistic fallacy, however, I disagree with Moore's description of moral properties as some kind of 'simple, unanalysable properties' which are wholly distinct from other natural properties. This is because this conception makes moral properties and moral facts to be some kind of queer and mysterious entities[36]. Moral facts, in my own understanding, are not entities. They are facts about how things are in all possible worlds. This fact justifies the truth of a moral judgment if and only if there is no possible moral world where it does not exist. What I accept from Moore and Mackie is that moral facts are not observable or discoverable by empirical investigation; they are, however, not entities, either mysterious or queer. They are facts in any possible moral world, which are appealed to in moral discourses.

Some possible problems that could be raised against my understanding of moral facts are: first, it could be argued that my account does not establish the ontology of moral facts. Unlike physical facts to which we can identify and observe, are moral facts identifiable or locatable in all possible moral worlds? This view rests wholly on ontological naturalism, which holds that "only natural objects, kinds and properties are real."[37] If this were correct, possible moral worlds would have to be physical or actual worlds, and moral facts would have to exist physically in such worlds. However, given our understanding of possible moral worlds as the way moral discourse could have been or a possible moral state of affair, possible moral worlds are not the same as physical worlds; hence, moral facts in these worlds are not physical facts; they are facts that exist in such possible worlds.

Another problem is that suppose it is conceded that moral facts are some kind of facts that exist in possible moral worlds, the question is how would this help the case of naturalizing ethics? In other worlds, since moral facts do not obey natural laws and principles, then ethics could not be declared a natural discipline. In response, I wish to argue that the project of naturalizing ethics needs not follow the way of ontological naturalism; I think it could be modeled

[36] This is the same sense in which moral facts have been described by J.L Mackie who noted that there were objective moral values (moral facts), then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we wee aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different form our ordinary ways of knowing everything else". See Mackie J.L. (1988: 115).

[37] Kim Jaegwon and Ernest Sosa. (1985: 343).

towards methodological naturalism. The methodological naturalists do not claim parity among all disciplines, they simply hold that "the best methods of inquiry in the social sciences or philosophy are, or are to be modeled on, those of the natural sciences."[38] In this respect, it is not essential that in modeling an inquiry on philosophy on the method of the natural sciences that all the apparatus used in one must be of the same kind in the other.

What I am trying to establish is that moral facts are facts of moral discourse, which obtain in every possible moral world. True moral judgments correspond to these facts in order to ascertain their truth or falsity. This method establishing the truth or otherwise of moral judgments is modeled on the method of testing theories in natural worlds. However, while in the natural sciences, scientists rely on their own kind of facts (natural) to confirm their theories, ethicists rely on their own kind of facts: moral facts, which exist in every possible moral world, to confirm their moral judgments. In order to confirm a natural judgment or theory, the natural or actual world is observed in order to establish the presence of natural facts, once these are discovered, the judgment is confirmed. In the same way, moral philosopher search through the possible moral worlds to establish that the fact of the moral judgment is present, once this is established, the moral judgment is confirmed.

5. *Conclusion*

The crux of one of Quine's argument in "On the Nature of Moral Values" that I addressed in this paper is that ethics does not belong to the same class of naturalism to which ontology and epistemology have been admitted. The main reason for denying ethics membership of naturalism, according to Quine, is that ethics as compared to science, is methodologically infirmed. This is because there are no moral facts in the world to which moral judgments correspond, through which moral judgments could be confirmed. Having explicated this problem, I attempted to show that, though not in the same natural or actual world, moral facts exist in every possible moral world. Once there is no possible world in which the judgment is contradicted, then the truth of the moral judgment is a moral fact. It is this moral fact that moral judgments correspond to in order to confirm their truth or falsity. Since, this is the method at play in

[38] Kim Jaegwon, et.al., (1985: 343)

science, I, therefore, think that Quine's charge that ethics is methodologically infirm can be challenged.

References

Flanagan O. J. 1982 "Quinean Ethics". *Ethics*. Vol. 93.
Gibson Roger F., "Flanagan on Quinean Ethics" (an unpublished version)
Hahn, L.E and Schilpp, P.A. 1986. (eds.) *The Philosophy of W.V.Quine*. La Salle, Illinois: The Library of Living Philosophers. Vol. XVIII.
Jaegwon Kim and Sosa Ernest. 1985. *A Companion to Metaphysics*, Oxford: Basil Blackwell.
Kirk, G.S., Raven, J.E., and Schofield, M. 1983. *The Pre-Socratic Philosophers: A Critical History with a selection of Texts*. Second edition. Cambridge: Cambridge University Press.
Lewis, David. 1973. *Counterfactuals* Oxford: Blackwell Publishers.
Lewis, David. 1986. *On the Plurality of Worlds*, Oxford: Basil Blackwell.
Moore G.E. 1971. *Principia Ethica*. New York
Plantinga Alvin. 1978. *The Nature of* Necessity. London: Oxford University Press.
Quine W.V. 1981. *Theories and Things*. Harvard: Harvard University Press
Quine, W.V.O. 1969. "Ontological Relativity" in *Ontological Relativity and other Essays* New York: Columbia University Press.
Rottschaefer, W.A. 1999 "Moral Learning and Moral Realism: How Empirical Psychology Illuminates Issues in Moral Ontology", *Behaviour and Philosophy*. Vol. 5.
Sayre-McCord Geoffrey. 1988 (ed.) *Essays on Moral Realism*. London: Cornell University Press.
Shirk Evelyn. 1965. *The Ethical Dimension*, New York: Appleton-Century Crofts.
Vlastos Gregory. 1946. "Ethics and Physics in Democritus" *The Philosophical Review*. Vol. 55. No. 1
G.F. Schueler. 1995. review of 'May, Larry; Friedman, Marilyn; and Clark, Andy, eds. *Mind and Morals: Essays on Ethics and Cognitive Science*. Cambridge, Mass.: MIT Press, 1996, pp. 315' in *Ethics,* Vol.107, No. 2, Jan.

Virginia Held, 2002. "Moral Subjects: The Natural and the Normative", *Proceedings and Addresses of the American Philosophical Association*, Vol. 76, No.2, November.

Etica & Politica / Ethics & Politics, X, 2008, 2, pp. 274-296

Identità personale, preferenze, narratività

Pierpaolo Marrone
Dipartimento di Filosofia
Università di Trieste
marrone@units.it

Abstract
The paper explores various conceptual formats that are designed to help us organize our thoughts concerning what we do, expecially personal identity, intertemporal preferences, and maximization. I discuss some suggestions from Parfit on personal identity and from Hare on interpersonal preference.

1. Indubbiamente, quando agisci hai intenzionalmente di mira il raggiungimento di un qualche obiettivo. Questo obiettivo sarà sempre specifico e singolare, anche se tu difficilmente ti accontenteresti di limitare le tue prospettive a un risultato particolare. In fin dei conti, la tua vita, come quella di chiunque altro, proprio perché è tua, soddisfa alcuni requisiti minimi di integrità personale e non ti appare come una rapsodia recitata da un attore diverso ad ogni istante, bensì come qualcosa che cerchi di comporre e di riconoscere come un'unità.

Gli esiti di questa operazione sono per lo più imperfetti, ma ciò che qui importa non è tanto la soddisfacibilità del risultato e nemmeno la sua praticabilità, quanto una certa costrizione, per non dire necessità, che ci guida a questo sforzo di coerenza narrativa, forse sempre incompiuta. Tutto questo vale per i tuoi atti intenzionali per lo meno, che per quanti pochi siano, sono quelli che ti vedono effettivamente agire come agente in prima persona. Non sto dicendo che siano solo o soprattutto gli atti intenzionali che danno sapore alla tua vita. Molto spesso sono le cose totalmente inattese a indirizzare nel bene e nel male le nostre vite e a dare loro il sapore specifico che hanno, e non invece i nostri atti intenzionali. Rimane però il fatto che quando agisci intenzionalmente, lo fai nella presunzione di essere proprio tu ad agire e non qualcun altro.

Le azioni intenzionali sono spesso compiute sulla base di quella che tu ritieni essere la migliore informazione disponibile al momento. Puoi naturalmente essere in errore e sbagliarti anche grossolanamente, ma è importante distinguere questa classe di azioni intenzionali da un'altra classe di azioni intenzionali, quelle dove decidi di ignorare la migliore informazione disponibile. Fatti salvi alcuni casi speciali, quando decidi di ignorare la migliore in-

formazione disponibile agisci in base a una forma di incontinenza. Ti accendi la sigaretta anche se hai un'informazione sufficientemente dettagliata dei danni che il fumo provoca. Una signora isterica ti passa platealmente davanti alla fila del supermercato. Reagisci in maniera fisicamente aggressiva anche se sai che la tua azione potrà avere conseguenze spiacevoli, ad esempio quelle contemplate dal codice penale o da un marito particolarmente prestante. Hai quindi agito in maniera sicuramente intenzionale, ma tuttavia con modalità incontinenti, le quali fanno sì che le tue azioni devono essere catalogate sotto il segno dell'irrazionalità. Tu stesso, se potessi fare una ripresa alla moviola della tua azione incontinente riusciresti ad indicare con una buona approssimazione il punto in cui le cose hanno cominciato ad andare per il verso sbagliato. È un segno questo alquanto dirimente del fatto che, secondo la tua stessa opinione, le cose non sarebbero dovute andare in quel modo, ossia del fatto che tu ritieni effettivamente di aver agito irrazionalmente e come non avresti dovuto agire sulla base del miglior giudizio disponibile per te. Si è stabilito quindi che le tue azioni intenzionali si possono raggruppare in almeno due categorie: quelle razionali e quelle irrazionali e incontinenti. A meno che non sia patologicamente portato ad azioni incontinenti che ti creano danni ripetuti, sono quelle che appartengono al primo insieme che costituiscono numericamente la maggior parte delle tue azioni intenzionali.

Quando tu agisci intenzionalmente e razionalmente che cosa ti prefiggi in effetti di ottenere? Al di là del risultato specifico, esiste uno scopo generale che è intenzionato dalla tua azione, ossia una sorta di oggetto ideale o una sorta di schema ricorrente, dotato di qualità specifiche che ricorrendo in atti appartenenti a quella determinata intenzionalità razionale li rende categorizzabili in un unico insieme? Io penso che le cose stiano effettivamente in questo modo, ma ritengo anche che l'analisi degli atti intenzionali razionali possa gettare una qualche luce non banale sul problema dell'utilità e su quello dell'intertemporalità, ossia su quella narratività che parrebbe essere un tratto costante della nostra descrizione come agenti morali.

Dunque: tu hai agito intenzionalmente e razionalmente. Che cosa hai fatto? Hai agito in maniera tale da soddisfare un qualche aspetto della tua volontà, ossia hai agito nella convinzione che con la tua azione avresti realizzato nella maniera migliore e nelle migliori condizioni conoscitive disponibili al momento per te un tuo desiderio. Tutto questo si può sintetizzare nella proposizione: (1) "agisco razionalmente quando faccio ciò che realizza un mio desiderio presente". Si noti che, sulla base di quanto finora si è detto, sarebbe pleonastico dire che agisco anche intenzionalmente. Se agisci razionalmente (ossia in maniera giustificata) allora agisci anche intenzionalmente. Non vale

invece la conversa, ossia, come si è visto, non è sufficiente, sebbene sia necessario, agire intenzionalmente per agire razionalmente.

2. Il problema è che (1) copre condizioni di razionalità che potrebbero apparirci troppo modeste, sulla base di quella narratività che la maggior parte di noi considera una caratteristica importante e forse essenziale della nostra autocomprensione come soggetti morali. Una soddisfazione di un tuo desiderio presente non è affatto detto sia ciò che tu razionalmente dovresti perseguire, ossia non è affatto detto che ciò che tu ora desideri sia anche ciò che ti darà la soddisfazione più intensa in un futuro che dovresti tenere in conto. Puoi desiderare intensamente al momento attuale di possedere una Porsche cabrio e, ammesso rientri nelle tue disponibilità finanziarie, andare dal concessionario a comprarla. Potrebbe però essere che la soddisfazione che evidentemente tu provi per il possesso di una bella macchina sia meglio soddisfatta da un altro modello che magari è maggiormente durevole, più semplice da guidare, meno impegnativo nella manutenzione e così via. In altre parole, è possibile che quello che tu desideri al momento attuale rappresenti su una scala una intensità massima, relativamente a un gruppo di desideri omogenei, ma non sia, invece, ciò che ti procurerà nel futuro la soddisfazione maggiore. Per questo motivo, se ti ritieni in grado di fornirti un'opinione razionale sul perché hai deciso di fare quello che stai per fare, dovresti compiere quell'azione che pensi massimizzerà la tua soddisfazione relativamente al desiderio che hai anche in un segmento temporale maggiormente esteso del presente.

Se (2) "per un agente agire razionalmente significa fare ciò che crede massimizzerà la soddisfazione di un suo desiderio attuale", allora noi introduciamo una dimensione temporale nella pianificazione delle tue azioni – ossia nella semplice tua volontà di essere razionale nell'azione – che in (1) non era precedentemente così chiara. Ma il passaggio da (1) a (2) è tutt'altro che innocente, dal momento che la dimensione temporale può diventare del tutto qualificante nella tua percezione di te stesso come agente razionale. Infatti, quali ragioni tu pensi di poter convincentemente esibire per dover preferire un determinato momento della tua vita anziché un altro? Perché privilegiare *t1* a *t2*, domani a dopodomani al 19 maggio del 2015? Che cosa possiede di così speciale domani rispetto ad innumerevoli altri momenti della tua vita futura?

Notoriamente, tanto Machiavelli quanto Hume erano profondamente pessimisti sulle capacità previsionali dell'essere umano a partire da una constatazione gnoseologica, che per loro si imponeva con la forza dell'evidenza empirica: noi siamo troppo ancorati a tutto ciò che percepiamo come presen-

te e prossimo per essere capaci di svolgere piani che vadano oltre la vicinanza temporale e spaziale. La maggior parte di noi è costretta nella gabbia del presente, e solo alcuni, talvolta – il buon politico, ad esempio, secondo Machiavelli, che deve però essere sorretto tanto dalla virtù quanto dalla fortuna - , ne possono evadere, per lo più momentaneamente. Tuttavia, le mie preoccupazioni non sono qui essere di natura descrittiva. Mi interessa piuttosto il versante normativo della questione e in quale modo questo aspetto si intrecci con quello della inevitabile temporalità dell'azione razionale. Un primo intreccio potrebbe proprio essere questo: non ci sono ragioni né a priori né universali e forse nemmeno generali per preferire un momento del futuro a un altro. Ovviamente, ci possono essere molte ragioni particolari perché nello specifico corso di azione che tu intraprendi queste ragioni vengano preferite. Se sei uno speculatore che opera in borsa, agirai diversamente a seconda che tu abbia una strategia rialzista o ribassista, nel senso preciso che valuterai in maniera molto diversa segmenti temporali diversi. Se sei una giovane donna in carriera potrebbe non essere indifferente il fatto che tu scopra di essere incinta. Gli esempi possono essere moltiplicati *ad libitum*, ma riguardano sempre occasioni specifiche di azione per le quali si adducono sempre ragioni particolari e non la struttura dell'azione razionale. All'interno di questa struttura sembra ragionevole non accordare preferenze particolari a segmenti temporali specifici. In questo senso, (3) "se tu agisci razionalmente farai proprio ciò che ritieni massimizzerà la soddisfazione di quei desideri compatibili con la tua razionalità, che pensi in qualche modo si estenderanno al corso della tua esistenza".

3. I tuoi stessi desideri presenti si presenteranno come degni di soddisfazione solo se saranno razionalmente compatibili con (3). Questo può causare, come si vedrà, dei problemi, perché quando si sostiene che il segmento temporale dove accade che i tuoi desideri siano soddisfatti non ha importanza, si sostiene che la narratività che costituisce la tua vita etica, narratività che naturalmente non può non essere temporale, prescinde da una determinata sequenza di segmenti temporali, ossia non può razionalmente farsi carico di preferire una data sequenza a un'altra. Questo in effetti pare essere alquanto controintuitivo e si cercherà di mostrare in queste pagine come il passaggio a (3) non sia affatto necessario.

Ammettiamo, tuttavia, per il momento che si tratti di una prospettiva invece plausibile per un agente razionale. Se lo facciamo non ci sorprenderà più di tanto veder sostenuta la posizione di chi pensa che (4) "non soltanto tu non dovresti prendere in considerazione la temporalità nella soddisfazione

dei tuoi desideri, ma non dovresti nemmeno a rigore preoccuparti di chi siano questi desideri". È questa la posizione di Parfit, per il quale l'identità personale, fondata sull'unità della coscienza, non è rilevante.[1] In effetti, sia alcuni degli esperimenti mentali immaginati da Parfit, sia descrizioni cliniche di soggetti affetti da lesioni cerebrali post-traumatiche sostengono la visione di Parfit dell'unità della coscienza come un'illusione. Non mi addentrerò in questo aspetto della filosofia di Parfit, anche se avrò modo di discutere alcune sue posizioni sulla neutralità temporale. Quello che mi preme invece suggerire con la presente discussione è che il passaggio a (3) non è inevitabile e, se questo sarà ritenuto ragionevole, allora forse anche il passaggio a (4) sarà ritenuto meno attraente. Naturalmente ci si potrà chiedere quale sia mai l'elemento attraente in (4). Penso sia piuttosto semplice dirlo: (4) aggiunge qualcosa di importante rispetto a (3) sul problema della neutralità. Nei nostri momenti migliori potremmo essere molto attratti dal pensare che è la struttura stessa dell'azione razionale a richiedere che tanto il tempo quanto l'identità personale dell'agente siano messi tra parentesi.

C'è un'accusa ricorrente che viene mossa all'utilitarismo e che è stata resa celebre da J. Rawls, da A. Sen e B. Williams,[2] ossia quella di non tenere in sufficiente conto dell'identità personale degli agenti. A parte il fatto che alcuni potrebbero non ritenerla affatto un'accusa e una carenza teorica, lo stesso potrebbe essere detto del contrattualismo o del kantismo. Lo si potrebbe dire di larghissima parte del codice civile e penale, eppure in questi casi nessuno pensa si tratti di una manchevolezza. Per quale motivo? Io credo perché dove è necessario che siano formulate regole di condotta e di limitazione dei comportamenti la prospettiva generale dovrebbe essere quella della clausola lockiana, ossia del non recare danno ad altri. La generalità della norma sembra escludere la narratività, almeno a un primo livello, ma non è invece così ovvio che la stessa esclusione debba essere richiesta quando parliamo di come sarebbe razionale che noi agiamo. In questo caso ne va della nostra capacità di concepirci, ognuno di noi, come centro intenzionale di azione.

Penso che questo ultimo punto non sia affatto non controverso, ma possa essere difeso con delle argomentazioni efficaci. Si prenda, ad esempio, la ben nota posizione che i desideri non esauriscono affatto il campo della motivazione. Si tratta della posizione kantiana, la quale è stata a più riprese soste-

[1] D. Parfit, *Ragione e persone* (1984), Milano, Il Saggiatore, 1989.
[2] A. Sen A., & B. Williams B. (a cura di), *Utilitarismo e oltre*, (1982), Milano, Il Saggiatore, 1984.

nuta anche recentemente.[3] Secondo la posizione kantiana si può essere motivati all'azione senza che intervengano desideri. Anzi, la motivazione ad agire non basata sui desideri rappresenta tanto l'agire etico quanto l'agire razionale nella loro purezza. Kant, come è noto, era talmente convinto della cosa da ritenere l'espressione 'ragion pura pratica' un puro pleonasma rispetto a 'ragion pratica'. L'azione motivata da desideri è impura in quanto vi entrano come fattori causali le inclinazioni personali. Le inclinazioni personali sono soggettive e non permettono di rintracciare quella struttura motivazionale universale dell'azione che è il segno della moralità; al contrario, quando la motivazione ad agire è costituita da un'obbligazione, i desideri non vi hanno parte. L'azione etica è universale precisamente perché motivata dal dovere morale che è universale e perciò impersonale.

Questa posizione rifiuta radicalmente la plausibilità di (1) ed ha avuto grande fortuna, ma a me sembra che si avvolga in aporie inestricabili e che non resista a una confutazione del genere seguente. Abbiamo visto che agire razionalmente significa agire intenzionalmente, sebbene non valga la conversa. Se agisci intenzionalmente ciò significa che alcuni tuoi atti mentali (credenze e progetti, e così via) hanno un potere causale sulla tua azione, ad esempio causano alcuni movimenti del tuo corpo che costituiranno proprio quell'azione che intendevi svolgere. Tutto questo potrebbe essere descritto anche in una maniera leggermente diversa e egualmente plausibile. I tuoi stati mentali, infatti, che influenzano anche la tua azione in maniera tale che questa si presenti come intenzionale, che cosa altro sono se non un desiderio di compiere proprio quell'azione? Sembra perciò difficile concepire che esistano ragioni che non diano luogo a desideri di compiere qualcosa. Se questo è plausibile, allora questa parte della filosofia morale di Kant deve essere rigettata. Infatti, l'azione intenzionale deve essere spiegabile, per il fatto stesso che è intenzionale, per mezzo di una relazione causale tra credenze, progetti e così via del soggetto. Questa spiegazione non è altro che una forma di causalità e, dal momento che credenze, progetti e così via che danno luogo a un'azione, danno luogo anche al corrispondente desiderio, sono i desideri ad avere forza motivazionale assieme alle credenze, ai progetti e così via che li generano.

Ciò che vale per l'azione intenzionale in genere, vale anche per quella razionale. La differenza più evidente è che nell'azione razionale dovrebbe essere l'informazione disponibile ad accordarsi con il desiderio. In altre parole, è la volontà di tener conto della migliore informazione disponibile che è disposi-

[3] O. O'Neill, *Acting on Principle*, New York: Columbia University Press, 1975; C.M. Korsgaard, *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press 1996.

zionale rispetto al pensiero proposizionale che fa di un desiderio anche un desiderio razionale. Esemplifichiamo. Io posso avere un intenso desiderio di diventare molto ricco, a partire da una situazione che rende questo desiderio estremamente improbabile, per la mancanza di capitali iniziali, di doti speculative, e di molte altre cose ancora. Se io acquisto un biglietto della lotteria, mi comporto irrazionalmente rispetto al mio desiderio e alle mie condizioni attuali? Dipende da quali altre informazioni io ho disponibili. Se io so che le mie probabilità di vincere una somma enorme, realizzando il punteggio massimo, sono all'incirca di una su seicentoventi milioni, e non ho altre chances ragionevoli di arricchirmi, e, allo stesso tempo, il costo della giocata ha un'influenza minima sul mio reddito, allora l'azione che si conclude nell'acquisto del biglietto della lotteria è generata da motivazioni e da un desiderio razionali. Se, viceversa, io non ho nemmeno la forza di recarmi a fare la giocata e preferisco fantasticare ad occhi aperti, non si può nemmeno dire che abbia delle intenzioni rispetto al mio desiderio – per quanto di improbabile realizzazione –, perché non ne esiste una traduzione in un'azione e il mio desiderio di arricchirmi rimane completamente vuoto. Rimarrebbe irrazionale anche se io dovessi giocare con la convinzione che le probabilità siano a mio favore per una qualche ragione.

Da questo non segue che due soggetti con un eguale bagaglio di informazioni avranno anche gli stessi desideri razionali. L'eguaglianza epistemica nelle conoscenze non porta nell'azione all'eguaglianza dei risultati, poiché, ad esempio, deve essere dato il peso adeguato a una eventuale differente propensione al rischio dei soggetti. Tuttavia, una volta accertatala, si genereranno insiemi differenti di desideri razionali, che potrebbero essere simili, se simile è la propensione al rischio dei soggetti. In altre parole, il potere causale di un identico bagaglio epistemico è una faccenda che deve essere risolta empiricamente, mentre rimane del tutto plausibile a priori che credenze, opinioni, progetti e così via siano necessariamente intrecciati a desideri che possono essere compatibili con questi e perciò razionali. In sintesi, sono questi i motivi per cui l'etica ha a che fare principalmente con desideri e deve essere rigettata la prospettiva etica kantiana che li concepisce come area dell'antropologia pragmatica.

4. Si è visto che il requisito temporale che si è enunciato in (1) presenta delle difficoltà che inducono a pensare che lo schema generale dell'azione razionale debba avere caratteristiche di marcata neutralità temporale. La neutralità temporale richiede, tra le altre cose, che l'agente razionale individui un desiderio che sia dominante intertemporalmente, non solo nella prospettiva di

evitare errori di calcolo morale, ma anche per evitare i paradossi dell'azione akratica. Parfit si è occupato estesamente delle difficoltà legate alla dimensione temporale delle varie teorie della razionalità pratica. Per illustrare i dilemmi intertemporali Parfit introduce la teoria degli obiettivi attuali (*present aim theory*) o P. "Supponiamo che in un dilemma del prigioniero il mio obiettivo sia quello di realizzare l'esito migliore per me. Secondo P in questo caso è razionale la scelta che arreca un beneficio a chi la compie. Se il mio obiettivo è quello di arrecare benefici agli altri o di superare il test kantiano, a essere razionale è la scelta altruistica. Se il mio obiettivo è di fare quello che fanno gli altri – magari perché non voglio essere un *free rider* – quale sia la scelta razionale è dubbio. Tutto dipende da quel che credo facciano gli altri".[4] Questa teoria è rigettata da Parfit, poiché P può essere in contrasto con i miei obiettivi di più lungo periodo. Il punto non è, tuttavia, solo questo. Parfit, infatti, segnala un contrasto ulteriore e maggiormente rilevante, vale a dire il contrasto tra P e la *teoria dell'interesse personale* o S, che è anch'essa una teoria della razionalità pratica. "S assegna a ciascuna persona questo obiettivo: conseguire quelli che per lei sarebbero gli esiti migliori e che consentirebbero alla sua vita di andare nel miglior modo possibile".[5] Posto questo obiettivo è lecito chiedere che cosa o quali condizioni lo realizzino. Parfit individua tre teorie che vi rispondono in maniera adeguata: a) l'edonismo; b) l'appagamento dei desideri; c) l'oggettivismo dei valori. "Tutte queste teorie sostengono inoltre che, nel decidere che cosa sia meglio per qualcuno, si dovrebbe assegnare lo stesso peso a tutte le parti del suo futuro. Gli eventi futuri possono essere meno prevedibili; e un evento prevedibile dovrebbe contare di meno se è meno probabile che accada. Non dovrebbe, invece, contare di meno per il solo fatto che, se accadrà, accadrà più tardi".[6] Questo requisito di neutralità temporale riguarda il carattere di razionalità pratica di cui è investita (S), ossia la sua portata deliberativa. In questo senso, è possibile riformulare (S) con "(S1) Per ogni persona c'è un fine ultimo sommamente razionale: che la sua vita sia, per lei, la migliore possibile".[7] Che il fine razionale ultimo sia la soddisfazione dei desideri dell'agente è un requisito il quale, unito a quello di neutralità temporale, porta a sostenere che l'agente dovrebbe fare ciò che comporta il soddisfacimento dei suoi desideri avendo come punto di riferimento l'intera sua vita. Si tratta di un requisito impegnativo, ma che riguarda unicamente la neutralità temporale non quella personale. Natural-

[4] D. Parfit, *Ragioni e persone*, cit., p. 119.
[5] D. Parfit, *Ragioni e persone*, cit., p. 6.
[6] D. Parfit, *Ragioni e persone*, cit., p. 6.
[7] D. Parfit, *Ragioni e persone*, cit., p. 7.

mente, c'è un contrasto tra (1) e (S1), dal momento che (1) si limita ad affermare che l'agente dovrebbe massimizzare i desideri che ha al tempo *t*. Sembra perciò rispondere all'evidenza che mentre (S1) è temporalmente neutrale, (1) invece non lo è.

Questo contrasto è realmente così forte? Non sembrerebbe, al contrario di quanto si sarebbe indotti di primo acchito a pensare. Quando si prescrive la massimizzazione dei desideri che attualmente si ha, si prescrive anche di non perseguire la massimizzazione dei desideri che si avranno presumibilmente nel corso della propria vita? Il conflitto tra (1) e (S1) è, suggerisco, un conflitto contingente, ma non necessario. È molto spesso del tutto razionale pensare che i desideri che l'agente cerca di massimizzare nel tempo *t*, potrebbero, almeno alcuni di loro, rientrare in uno specifico sottoinsieme. Questo sottoinsieme comprenderebbe desideri reiterati nel corso del tempo. Lasciando da parte i desideri akratici, è possibile ipotizzare che ci siano desideri che, massimizzati nel presente, estendono il loro potere causale nel futuro, rispondendo così esattamente al requisito richiesto da (S1). Tuttavia, c'è anche un'altra ragione per pensare che il conflitto tra (S1) e (1) non sia un conflitto tra due versioni profondamente differenti della razionalità pratica. Si prenda la notazione precedente sulla motivazione. Se è vero che pensieri, credenze, progetti hanno un potere causale sull'azione, in maniera tale che è ragionevole che l'agente abbia il desiderio di agire in conformità a questi stessi pensieri, credenze, progetti che attualmente ha, allora anche quando l'agente compie qualcosa in conformità a quanto richiesto da (S1), agisce sulla base di un desiderio presente massimizzandolo. Si noti però che può essere avanzata anche un'altra considerazione di un certo interessere. Si tratta di questo: poiché agisco in base a pensieri, credenze, progetti che hanno un potere causale sui desideri, esiste un desiderio soverchiante, che forse sarebbe meglio chiamare meta-desiderio, di agire in conformità ai propri progetti, credenze, progetti. Questo meta-desiderio non può essere negato se non generando contraddizione. Mi spiego con un esempio. Ammettiamo che tu sia preda in periodo particolarmente difficile della tua vita di un determinato automatismo di pensiero, che genera credenze errate e sofferenze. Tutto ciò ha un potere causale sul corso delle tue azioni, in primo luogo perché queste credenze necessariamente generano un desiderio di agire in conformità ad esse. Ammettiamo ora che tu voglia liberarti da quello che percepisci come un meccanismo generale perverso, in maniera tale che i tuoi pensieri, credenze, progetti precisamente non generino il desiderio di agire conformemente ad essi. L'unica risorsa a te accessibile pare essere che tu ti formi altre credenze, pensieri, progetti, ossia che tu intraprenda un processo di revisione critica delle tue credenze attuali al fine di sostituirle con altre credenze. Non vi è maniera di sfuggire a questa

connessione tra credenze e desiderio soverchiante. Anche se tu dovessi abbracciare una soluzione estrema, quale porre fine alla tua vita cosciente, non vi saresti in effetti sfuggito, perché non avresti fatto altro che confermarlo con il tuo gesto. Dopo tale atto estremo non potresti più essere semplicemente definito un soggetto agente.

5. Il legame tra pensieri, credenze, progetti e il desiderio di agire in conformità a questi non può, cioè, essere negato se non sostituendo a determinati progetti, credenze, pensieri altri specifici progetti, credenze, pensieri. Io penso che siamo in presenza di una struttura trascendentale, nel senso che il legame così individuato tra pensiero e desiderio costituisce una struttura necessaria per l'azione. Questa notazione può essere interessante proprio per il requisito della neutralità temporale richiesto da Parfit. Questa struttura dell'azione è temporalmente neutrale, anche se non forse nel senso normativo che vorrebbe Parfit. Per Parfit, infatti, tu non hai nessuna ragione a priori per preferire un determinato segmento temporale a un altro; quindi, non dovresti farlo. Nel caso che illustravo siamo in presenza di qualcosa che è invece strutturale e fondamentale. Il legame pensiero-desiderio è temporalmente neutro sia che tu lo decida sia che tu non lo decida. È semplicemente una descrizione di qualcosa che accade sempre nell'azione. In questo senso, penso che (S1) debba essere incorporata in (1), ossia (S1) deve essere considerata un caso specifico di (1). Si potrebbe sostenere che, in realtà, le cose stanno in maniera esattamente contraria a quanto si sta affermando, e che è (S1) a incorporare (1), sebbene in una versione leggermente modificata di (1). Il suggerimento è stato, in effetti, ancora una volta, avanzato da Parfit. "Secondo tale teoria certi tipi di obiettivi, pur sopravvivendo a tale processo di deliberazione, sono intrinsecamente irrazionali e non forniscono ragioni per l'azione. Ciascuna persona ha più ragione di fare ciò che meglio realizzerà, degli obiettivi attuali, quelli che non sono irrazionali. È la *teoria critica degli obiettivi attuali*".[8] Qui si presenta un problema, ossia quello di determinare se esistano effettivamente desideri che siano intrinsecamente irrazionali. 'Intrinsecamente irrazionali' significa 'irrazionali a prescindere dal contesto'. Il contesto comprende (a) sia le informazioni migliori disponibili all'agente, (b) sia lo sfondo precedente delle sue valutazioni pregresse, (c) sia il suo sistema di valori, (d) sia la capacità di rappresentarsi in maniera adeguata i tre elementi precedenti. Trovo effettivamente difficile immaginare un esempio che possa sostenere l'esistenza non contestuale di desideri intrinsecamente irrazionali. Certo, tu

[8] D. Parfit, *Ragioni e persone*, cit., p. 122.

potresti voler ripetere le imprese di Alessandro Magno o di qualche altro grande condottiero del passato perché pensi di esserne un discendente in linea diretta ed averne così acquisito le abilità strategico-militari incorporandole nel tuo codice genetico. Il desiderio è irrazionale perché il contesto dal quale nasce ha a che fare con la patologia mentale. È molto facile escogitare innumerevoli altri esempi di questo genere ed il caso non è, perciò, particolarmente interessante.

Ammettiamo, tuttavia, che tu sia a conoscenza della struttura generale del dilemma del prigioniero e del risultato sub-ottimale che si ottiene scegliendo la strategia dominante. Ammettiamo che tu scelga ora effettivamente questa strategia. Siamo effettivamente in presenza degli elementi necessari per poter sostenere che tu hai fatto la tua scelta perché preda di un desiderio intrinsecamente irrazionale? Potresti essere stato mosso da considerazioni di sfondo che basterebbero a giustificare la tua azione e a salvaguardare, dal non essere intrinsecamente irrazionale, il tuo desiderio di compiere proprio quella scelta che ti conduce all'esito sub-ottimale. Qualora queste altre considerazioni non fossero presenti e tu semplicemente ti trovassi in una situazione descrivibile formalmente secondo lo schema del dilemma del prigioniero, allora è chiaro che basterebbe la consueta carenza informativa prevista dal dilemma stesso per non rendere la tua azione irrazionale, ma semplicemente aderente alla strategia dominante. Quello che voglio suggerire è che nella procedure deliberative non esistono condizioni oggettive indipendenti dagli agenti che consentano di etichettare a priori un tuo desiderio come 'intrinsecamente irrazionale'. Tutto questo non deve essere confuso con una posizione relativista rispetto ai valori, piuttosto si intende semplicemente affermare il valore strumentale dei desideri. Rispetto a questo aspetto strumentale i valori sono come degli oggetti ideali, e la loro idealità, ossia la loro relativa non contestualità, non esclude affatto un ordine gerarchico interno ai valori.

Si è visto come Parfit invochi la neutralità temporale come un elemento di superiorità di (S1) rispetto ad altre concezioni, ad esempio, a quella della razionalità deliberativa espressa da (1). Si è visto anche la concezione deliberativa e causale di (1) non supporta queste conclusioni di Parfit. È utile qualche ulteriore precisazione riguardo alla concezione della neutralità temporale che qui è in gioco. Infatti, io penso se ne possano distinguere per lo meno due: da una parte, c'è (a) "la nozione di neutralità temporale relativa ai desideri di colui che compie l'azione". Quello che la neutralità temporale prescrive è che i desideri dell'agente nel corso del tempo, per quanto possibile, siano conosciuti e, per quanto è possibile, ne sia conosciuta la loro intensità. La loro importanza dovrebbe essere proporzionale alla loro intensità e

non alla loro collocazione nel segmento temporale della vita dell'agente. Questa concezione ritiene sia importante unicamente l'intensità del desiderio e che questa sia la sola base per considerarlo come desiderio (relativamente) dominante per l'agente. Da un'altra parte, c'è una concezione ancora più esigente della neutralità temporale. Secondo quest'altra concezione, (b) "non ha alcuna rilevanza che un desiderio sia proprio di un agente oppure di un altro, se la loro unica differenza consiste in un differente posizionamento nel continuo temporale". In questo senso, se io ho la possibilità di soddisfare un mio desiderio al tempo *t1* che è identico al desiderio che un altro avrebbe al tempo *t2*, e io giustifico l'azione che realizza il mio desiderio sostenendo non tanto l'appartenenza di tale soddisfazione a una biografia, quanto la maggiore prossimità al presente, allora agisco in violazione della neutralità temporale. Se io invece giustificassi la mia scelta sostenendo la superiorità valoriale del mio desiderio, non avrei effettuato nessuna violazione della neutralità temporale: in definitiva, poiché si farebbe riferimento a due differenti strumentalità in vista del raggiungimento di due diversi ordini di valori, se ne potrebbe legittimamente concludere che siamo in presenza semplicemente di due desideri diversi.

6. Entrambe queste concezioni sono visioni egualitarie della soddisfazione dei desideri. La seconda concezione è però radicalmente egualitaria, dal momento che mette tra parentesi l'identità personale dell'agente. Questa seconda idea di neutralità temporale e interpersonale presenta molteplici problemi. Il principale tra questi a me sembra essere che ogni deliberazione rischierebbe di richiedere qualcosa di molto prossimo all'etica supererogatoria. Ci si allontanerebbe in tal modo in maniera irrimediabile da qualsiasi cosa possa essere considerata una descrizione plausibile della nostra personale esperienza morale. Inoltre, si appiattirebbe l'esperienza morale su una dimensione sorprendentemente deontologica e irrealisticamente esigente. Di questa interpretazione radicale della neutralità temporale e interpersonale, quindi, non mi occuperò oltre. La prima interpretazione mi sembra invece maggiormente interessante e suscettibile di analisi. Una volta che l'identità personale non viene più considerata come un tratto problematico, ciò che rimane e viene espresso dall'idea di neutralità temporale è semplicemente lo scopo dell'azione razionale, a mio modo di vedere. Tale scopo è la massimizzazione intertemporale della soddisfazione dei desideri dell'agente. Si tratta di una tesi plausibile e sostenibile? Cominciamo da una questione riguardante i desideri passati. Ogni agente è in grado di distinguere i desideri che ha al presente da quelli che ha avuto nel passato. Tra questi ultimi, ve ne sono altri che ha avuto e che

attualmente ha ancora, o in forme identiche o in forme comparabilmente simili a quelli del passato. L'agente dovrebbe tenere presenti questi desideri nel suo calcolo? Da un lato, sembrerebbe fin troppo ovvio sostenere che i desideri passati possono contribuire potentemente a formare le nostre preferenze attuali. Non sembrerebbe quindi esserci motivo per escluderli da una considerazione presente. Questo però ci condurrebbe direttamente a un paradosso. Il peso dei desideri è valutabile dal grado di soddisfazione che io posso assegnare loro. I desideri passati non possono per definizione essere soddisfatti. Si potrebbe argomentare che effettivamente non esiste una soddisfazione positiva dei desideri passati, ma che questo non significa che noi non possiamo collocarli lungo una scala di soddisfazione. Precisamente la loro collocazione in questa scala è dalla parte dei valori negativi. Questa, però, a me pare essere un'inutile complicazione. Il fatto che un desiderio non ci sia più non equivale alla sua frustrazione. Una frustrazione non è tanto una soddisfazione negativa, quanto una soddisfazione negata. Dal momento che soddisfare un desiderio significa pianificare le condizioni materiali che condurranno alla sua realizzazione, allora soltanto i desideri presenti e futuri possono avere soddisfazione, ma non quelli passati, che, quindi, a rigore non possono nemmeno essere frustrati.

Sembrerebbero considerazioni banali, ma non nel senso che riflettere sulle condizioni di realizzabilità dei desideri comporta anche una sorta di terapia cognitiva per l'agente. Un risultato di questa terapia, per così dire, è che il passato è irredimibile e non può essere scontato. Tu non puoi, quindi, avere obiettivi rispetto al tuo passato, poiché i desideri che avevi non ci sono più. Se tu pensi altrimenti sei chiaramente vittima di un errore cognitivo e il tuo modello deliberativo dovrebbe essere sottoposto a revisione, la quale, come si diceva, avrebbe anche un qualche valore terapeutico. Quello che conta è perciò il futuro. Ma rispetto al futuro quali desideri dovresti avere? Ognuno ne ha molti che possono essere raggruppati in insiemi diversi, ma per i nostri fini è utile distinguere soprattutto due gruppi. Il primo comprende quei desideri che, sulla base delle tue informazioni attuali, se realizzati, condurranno a massimizzare una qualche tua soddisfazione futura. Il secondo gruppo comprende un altro genere di desideri. Si tratta di quei desideri che in qualche modo l'agente ritiene di poter plasmare. Questi desideri giocano un ruolo molto importante nelle nostre vite, dal momento cha altro non sono che preferenze autoindotte. Di solito, le preferenze autoindotte – ad esempio, la decisione di intraprendere una determinata carriera, di intrecciare una relazione affettiva stabile, di sottoscrivere un mutuo ventennale per acquistare un appartamento – implicano una pianificazione del futuro. Non solo: implicano anche una diversa concezione della massimizzazione. Rispetto alle preferenze

autoindotte io devo nutrire la convinzione che la loro soddisfazione sia intertemporale. Ad esempio, devo pensare che se ho sottoscritto un mutuo ventennale a rate mensili superiori (ma non troppo) a quanto pagherei per un affitto, la mia soddisfazione sarà massimizzata ad ogni pagamento mensile, anche prima dell'estinzione del mutuo, perché ogni rata pagata mi approssima a quel risultato, anche se a questa soddisfazione devo aggiungere quella relativa all'approssimarsi al risultato finale. Si noti bene che queste preferenze non implicano affatto una qualche etica del sacrificio (che pure deve alimentarsi, soprattutto se non esclusivamente, di preferenze autoindotte), ma è una semplice descrizione di una modalità fondamentale dell'azione che tutti noi adottiamo. Una modalità di massimizzazione diretta di preferenze attuali può invece condurci spesso a risultati indesiderati. Infatti, potrebbe essere che un agente abbia al momento attuale, poniamo, un intenso desiderio di vendetta verso un'amante che lo ha tradito. Si trova anche nelle condizioni di poterla esercitare senza probabilmente subirne le conseguenze. Sarebbe tuttavia irrazionale per lui realizzare ciò che è al momento il desiderio dominante, dal momento che non è certo che dal suo atto non discendano conseguenze negative delle quali potrebbe pentirsi in futuro. Inoltre, poiché ognuno sa che i desideri molto violenti tenderanno ad affievolirsi nel corso del tempo piuttosto rapidamente, l'agente sa già ora che le sue preferenze molto probabilmente tenderanno a cambiare. Facciamo, però, un altro esempio che risulta essere maggiormente convincente. Immaginiamo che tu in assenza di gravi malattie fisiche invalidanti sia profondamente depresso e che tu nutra desideri autodistruttivi. Immaginiamo anche che a uno stadio della tua depressione questi divengano dominanti. Ti comporteresti irrazionalmente se tentassi di soddisfarli, perché è ragionevole pensare che, se i tuoi desideri depressivi non saranno più dominanti in futuro, il saldo netto della rimanente porzione della tua vita potrà essere positivo. In realtà, per fare della soddisfazione dei tuoi desideri autodistruttivi una scelta ampiamente irrazionale sarebbe sufficiente che il tuo saldo netto di soddisfazione per quanto ti rimane da vivere sia leggermente positivo. Infatti, se tu massimizzi la soddisfazione dei tuoi desideri autodistruttivi, alla fine non ci sarebbe nessun saldo positivo. Salvo circostanze molto specifiche (dolorose malattie terminali e atti di sacrificio supererogatori), spesso è irrazionale escludere che non ci possa essere un maggior saldo netto da distribuire lungo l'arco temporale che non sopravanzi un intenso desiderio dominante da soddisfare al presente.

7. Le ragioni estese nel corso del tempo sono particolarmente evidenti in due circostanze generali. Quelle relative all'apprendimento e quelle relative alla

disassuefazione da sostanze psicotrope. Merita esaminare qualche esempio in proposito. Immaginiamo che io sia un dirigente di medio livello di una multinazionale con forti interessi in Cina. Le mie prospettive future di carriera sono legate al fatto che apprenda in un tempo ragionevolmente breve il cinese mandarino. Ho perciò al momento attuale una motivazione ad apprendere il cinese mandarino, non soltanto perché la soddisfazione che potrei trarne adesso, se io lo sapessi già, è soverchiante, ma anche perché è plausibile pensare che lo sarà, poniamo, tra sei mesi. La motivazione a soddisfare il mio desiderio è perciò svincolata dal presente, nel senso che trae la sua forza (e certamente la sua difficoltà) dal non essere limitata all'attualità. Tra sei mesi è perciò probabile che la motivazione ad apprendere il cinese mandarino avrà conservato la sua validità intrinseca, nel senso preciso che tra sei mesi io potrò ancora sostenere che avevo le ragioni di pensare sei mesi fa che dopo sei mesi la motivazione ad un apprendimento probabilmente faticoso avrebbe conservato la sua capacità motivazionale. Sembrerebbe che lo stesso si verifichi con i migliori programmi di disassuefazione da droghe. Il consiglio di mantenersi puliti per almeno ventiquattro ore e di pensare di mantenersi puliti nelle successive ventiquattro quando queste si presenteranno e così via, va esattamente nella medesima direzione, poiché confida sul fatto che la forza motivazionale si mantenga per lo meno identica distribuendosi nel corso del tempo.

In nessuno di questi due esempi la capacità di offrire valutazioni sui propri stati futuri è indipendente dalla volontà di soddisfare desideri presenti. Accettare la positività di uno stato ed assegnargli un valore positivo implica, alla luce della discussione precedente, che si ha un desiderio attuale per quello stato, poiché le motivazioni non possono non essere sorrette dai desideri. Il risultato è che anche le motivazioni maggiormente legate a quelli che si ritengono essere dei valori di alto profilo assumono la loro forza se noi li desideriamo e se pensiamo che soddisfarli sia meglio che non soddisfarli. Dal momento che la soddisfazione di questi valori di alto profilo è tipicamente intertemporale, tu devi essere in grado già ora di concepire che, desiderando nel momento attuale la tua soddisfazione relativamente a questi, la desidererai anche nel futuro in maniera analoga. Disporre la soddisfazione in senso intertemporale rimanda alla capacità di sapere come saranno le proprie preferenze nel futuro, ossia rimanda al ruolo dell'immaginazione nell'azione. Per dirla con le parole di Hare "Qui incontriamo nuovamente un'intima relazione concettuale fra gli stati cognitivi, affettivi, e conativi, ma questi ultimi hanno per oggetto degli stati di cose ipotetici, non reali".[9] Questa dimensione ipote-

[9] R. Hare, *Il pensiero morale* (1981), Bologna, Il Mulino, 1989, p. 134.

tica è della più grande importanza per renderci comprensibili le esperienze che gli altri fanno e quelle che potremmo fare noi. In altre parole, è parte integrante dell'esperienza umana che io in futuro possa avere le preferenze di un altro. "Supponiamo che io dica: 'Sì, so esattamente come ti senti, ma non mi importerebbe affatto se ora qualcuno facesse lo stesso a me', non dimostrerei forse che in realtà non sapevo, e nemmeno pensavo che ci si sentiva *in quella maniera*?".[10] In maniera del tutto analoga, se ho delle preferenze per situazioni future, in certa misura devo essere in grado di replicarle al presente, immaginando come mi sentirei se fossi capace di soddisfarle nel segmento temporale in cui mi ora le sto collocando.

Per questo motivo Hare può sostenere che la seguente proposizione: " (1) io attualmente preferisco con intensità I che, se fossi in quella situazione, avvenga X anziché il contrario" unita a "(2) se fossi in quella situazione preferirei con intensità I che X avvenisse anziché il contrario"[11] esprimono una verità concettuale. Per quanto non siano affatto proposizioni equivalenti, se uso nel senso usuale il verbo 'sapere' non posso sostenere la verità di (2) a meno che non sottoscriva la verità di (1). Da questa affermazione Hare ne trae un'altra di estremo interesse, ossia l'idea che il termine 'io' non sia per intero un termine descrittivo, bensì inglobi anche una dimensione prescrittiva. "Identificandomi realmente o ipoteticamente con un'altra persona, io mi identifico con le sue prescrizioni".[12] Detto in altre parole, quella che è stata identificata come 'funzione narrativa' indispensabile a pensare la propria esperienza etica come propria, presuppone la capacità di riferire questa esperienza a un nome, ossia a una descrizione definita. Questo non significa affatto dare per risolto il problema dell'identità personale, che comporta molte questioni di ordine differente (gnoseologiche e metafisiche), ma più semplicemente indica la capacità di riferirsi a un nome. Questo spiega anche l'efficacia della deterrenza. "Se considerare la persona che viene punita come me stesso, implica avere un'avversione contro il fatto che egli venga punito pari alla mia avversione futura, questo spiega perché io eviti di commettere il reato per il quale egli viene punito".[13] Se io tengo conto delle mie attitudini cognitive, posso allora far assumere a quelli che sarebbero dei meri desideri futuri la forza di preferenze attuali.

Questa attitudine simpatetica verso preferenze altrui può essere illuminante per quanto riguarda le ragioni per cui io avrei dei buoni motivi per cu-

[10] R. Hare, *Il pensiero morale*, cit., p. 134.
[11] R. Hare, *Il pensiero morale*, cit., p. 136.
[12] R. Hare, *Il pensiero morale*, cit., p. 137.
[13] R. Hare, *Il pensiero morale*, cit., p. 136.

rarmi di preferenze che potrebbero essere mie. Hare sostiene che tutto questo non coinvolge il problema dell'identità personale. Io penso che occorra essere leggermente scettici sulla validità di questa affermazione, poiché credo sia inevitabile interrogarsi sulle ragioni che effettivamente io potrei avere per interessarmi delle mie preferenze future. L'unica risposta che io riesco a immaginare è una risposta basata su una soluzione specifica del problema dell'identità personale, e precisamente quella basata sulla continuità psicofisica nel corso del tempo. Del resto, è piuttosto facile ammettere che il sentimento di simpatia che io posso provare nei confronti di un altro e che mi permette di immaginare alcuni suoi ordini di preferenze, è un sentimento non originario, ma piuttosto derivato. Per quanto io possa essere simpatetico verso gli altri, lo sarò sempre di meno che nei confronti di me stesso. E se io sono simpatetico verso me stesso non lo sono tanto e soltanto verso l'io che adesso puntualmente sono, bensì piuttosto verso questo io in quanto si nutre delle aspirazioni e dei progetti che precisamente lo costituiscono in quanto è l'io che adesso è, ossia, perché riesco a immaginare una continuità di me stesso nel futuro come unità psicofisica. Non sto sostenendo che il fatto che il mio atto immaginativo sia efficace renda metafisicamente reale questa unità. Ci sono delle buone ragioni, anzi, per continuare a metterla in questione. Quanto sostengo, invece, è che la credenza sulla sussistenza relativa di tale unità è quella che rende ragionevole preoccuparsi delle nostre preferenze future ed anche è quanto fonda il nostro interesse simpatetico per le preferenze degli altri.

8. Mentre questa implicazione mi sembra del tutto corretta e necessaria alla prospettiva simpatetica che si è costretti ad assumere nella valutazione intertemporale delle preferenze, la medesima prospettiva rende anche dubbia la specifica prospettiva di Hare riguardo l'universalizzazione. Hare, infatti, ha ampiamente sostenuto che le nostre ragioni per agire moralmente sono chiaramente fondate perché rispondono a una determinata logica, quella dell'universalizzazione appunto. Se sei in grado di pensare che la tua azione debba essere compiuta da chiunque in circostanze simili alle tue, allora siamo nella sfera morale, altrimenti rimaniamo confinati alla sfera della prudenza personale, se non addirittura dell'egoismo. Tuttavia, il riferimento alla continuità psicofisica dell'agente che io sono non può palesemente essere universalizzabile. Sono io che immagino questa unità futura. Pretendere che non solo venga immaginata, ma anche assunta nei miei medesimi termini sarebbe un narcisismo che inclina verso la patologia. D'altra parte, l'imputazione di quella unità ad altri è operazione non meno dubbia. Tutti sappiamo che le

persone cambiano e non ci scandalizziamo affatto se in circostanze simili, ma posti in due segmenti temporali ragionevolmente distanti, operano scelte o avanzano ragioni per giustificare i propri atti che non sono simili. Questo non ci indirizza verso una eliminazione del pronome 'io' dal nostro lessico; anzi: ci sono molte buone ragioni per continuare a mantenerlo non soltanto dal punto di vista legale e giuspositivo della necessità dell'imputazione personale (che per altro comporta eccezioni nello stesso diritto, come è noto), ma anche per l'opportunità di mantenere tale 'centro focale' a fini di unità normativa. Certamente, siamo qui in presenza di una ambiguità. Da un lato, infatti, l'universalizzazione richiede l'eliminazione di tutte le variabili individuali che rendono sensato, narrativamente ed eticamente sensato, l'uso del pronome 'io', in senso specifico, ossia come *token*. Da un altro lato, secondo una interpretazione più stretta dell'universalizzazione, la motivazione di un'azione può essere universalizzabile a patto che in essa precisamente non compaiano riferimenti né a variabili né a costanti individuali. Questa versione dell'universalizzazione, che a me pare una versione particolarmente stringente di kantismo, ritiene che il riferimento all'io tanto come *token* quanto come *type*, renda l'atto di universalizzare la motivazione impuro. Infatti, io posso interpretare le preferenze di un soggetto A come accettabili da un soggetto B, senza che questo implichi in alcun modo che un terzo soggetto C debba farle proprie. Questo, in realtà non varrebbe nemmeno per B, se noi accettiamo l'idea che il pronome 'io' implichi una prescrizione, prescrizione in linea con una unità psicofisica nel corso del tempo. Infatti, né B né C sono continui con A; per quanto simili possano essere le loro preferenze, questo non significa che debbano adottare quelle di A per se stessi. Per quanto simili siano le loro motivazioni, questo non significa che debbano riconoscere un paradigma di generalità in A tanto forte da trasformarsi in universalità per ciascuno di loro. In definitiva, quello che l'interpretazione dell'io come termine descrittivo e prescrittivo richiede è una capacità di immedesimazione tale da poter immaginare che la soddisfazione delle preferenze di altri soggetti simili a lui per aspetti rilevanti, abbiano lo stesso peso della soddisfazione delle proprie preferenze. Per questo motivo Hare può scrivere. "Ora, i casi multilaterali presentano meno difficoltà di quanto sembrasse a prima vista. Infatti, anche in tali casi i conflitti interpersonali, per quanto complessi siano e per quante persone vi siano coinvolte, si ridurranno a conflitti intrapersonali, posto che si dia una completa conoscenza delle preferenze altrui".[14] In realtà, questa mossa è cruciale per l'idea che la logica interna del prescrittivismo conduca a una forma di utilitarismo. Ma se le cose stanno in questo mo-

[14] R. Hare, *Il pensiero morale*, cit., p. 152.

do, a me pare che siamo precisamente ricondotti a quella forma particolarmente esigente di universalizzazione di cui si diceva sopra. Il fatto è che l'identità numerica dovrebbe essere esclusa da una forma rigorosa di utilitarismo, poiché la fonte di legittimazione per la soddisfazione di alcune preferenze anziché di altre si ridurrebbe a una forma di pregiudizio, quel pregiudizio che ci fa usare il pronome 'io' al tempo stesso come un indicatore di una descrizione definita, che può quindi essere sostituita da un nome proprio, e come qualcosa che può essere assunto in linea di principio da chiunque altro, senza che vengano alterate le condizioni di prescrittibilità di una motivazione specifica, condizione senza dubbio estremamente esigente. Se queste osservazioni sono pertinenti se ne dovrebbe ricavare che la derivazione dell'utilitarismo dal prescrittivismo, che costituisce l'ambizione teorica principale di Hare, non è destinata ad essere soddisfatta, per lo meno in questi termini. Inoltre, possono essere avanzati dei dubbi sulla legge psicologica che sembra sottostare alla posizione prescrittivista. Questa legge o intuizione psicologica suona più o meno in questi termini: "se io giudico che esista un ente psicofisicamente continuo con me stesso, allora sarà razionale che desideri che le preferenze di tale ente siano soddisfatte in misura proporzionale alla loro intensità". In altre parole, la mia preoccupazione sulla soddisfazione delle preferenze di tale ente è garantita dalla sua continuità con me stesso. Ma ammettiamo che io vada a dormire e che per qualche sconosciuto processo atomico il mio corpo e la mia mente si dissocino per essere ricomposti quando suona la sveglia, per avvertirmi che devo alzarmi e prepararmi per andare a lezione, in una unità che è leggermente diversa da quella che avevo quando sono andato a dormire, leggermente diversa ma indistinguibile agli effetti macroscopici. Avrei davvero acquisito qualche motivo per non preoccuparmi della soddisfazione delle preferenze di questa entità leggermente discontinua con quella che esisteva quando sono andato a coricarmi? Noi sappiamo che nella realtà le cellule del nostro corpo vengono sostituite progressivamente. Non pensiamo che questo comporti una maggiore o minore preoccupazione per la soddisfazione delle preferenze di tale entità che muta. Per quale motivo, allora, le cose dovrebbero cambiare se il ricambio cellulare dovesse avvenire simultaneamente? Quello che vorrei suggerire è che sembra ovvio che l'identità personale debba far riferimento a una qualche forma di continuità spazio-temporale, poiché la sola continuità materiale non ne è né condizione sufficiente, ma nemmeno condizione necessaria. Preoccuparsi della soddisfazione delle preferenze di un sé successivo, cosa che normalmente accade quando si ritiene di essere in presenza di una qualche forma di continuità dei sé, non è causato dalla mera continuità, ma piuttosto dal fatto che questa forma sia una somiglianza prossima a un qualche sé passato. Esistono

molto probabilmente delle pressioni evolutive che fanno sì che noi, esseri dotati di autocoscienza, reagiamo in maniera così pronta e sensibile alla somiglianza dei nostri sé successivi. Dovessimo comportarci altrimenti, probabilmente non sopravvivremmo a lungo. Che questo sia un investimento emotivo o razionale o più probabilmente un intreccio tra i due non penso sia necessario e importante stabilirlo ai nostri fini, ma è questo investimento a rendere ragione del fatto che ci preoccupiamo maggiormente dei sé che ci sono prossimi anziché di quelli che ci sembrano estranei. Questa era stata anche l'intuizione di Hume, quando aveva rilevato che i meccanismi della simpatia non si estendono più di tanto nel tempo e nello spazio. Hume non pensava tanto alla simpatia nei confronti di sé successivi in quel contesto, quanto alla simpatia che ci è difficile estendere oltre la cerchia di coloro che ci sono più prossimi. Nei termini della presente discussione, tuttavia, la cosa potrebbe essere formulata nei seguenti termini: ci può essere somiglianza anche senza continuità. La legge psicologica che sembrerebbe stare alla base della razionalità a preoccuparsi della soddisfazione delle preferenze di sé ritenuti successivi in virtù della continuità spazio-temporale deve perciò essere abbandonata perché non necessaria, dal momento che ci può essere somiglianza senza che ci sia continuità. Ma deve essere abbandonata anche per un altro motivo, ossia perché ci può essere continuità senza che ci sia somiglianza. Mi rendo conto che mentre la somiglianza senza continuità può essere facilmente accettata, per quanto riguarda il secondo caso le cose sembrano, almeno intuitivamente, maggiormente controverse. Eppure casi di questo genere sono tutt'altro che rari. Immagina che tu sia una persona assolutamente media, rispettosa della legge, leale verso i tuoi amici e verso la tua famiglia. Per uno di quegli strani casi della vita che mai avresti contemplato, un giorno assumi una droga che ti costringe da subito a un'elevata dipendenza, facendoti divenire un tossicodipendente la cui principale preoccupazione è procurarsi la prossima dose, in spregio alla legge e violando ogni regola di lealtà verso i tuoi amici e verso la tua famiglia. Ammettiamo pure che tu sia stato costretto a diventare un tossicodipendente e che quindi tu non sia responsabile di ciò che sei diventato. Dal momento che prima di essere un tossicodipendente non avresti mai voluto diventarlo, nel senso preciso che non avresti riconosciuto le preferenze di un tossicodipendente come tue, non si può certo dire che il tuo sé anteriore avrebbe dovuto in qualche modo preoccuparsi della soddisfazione delle preferenze di un sé che mai avrebbe voluto essere.

Ammettiamo, tuttavia, che il presupposto della continuità psicofisica sia valido. Anche in questo caso non ne conseguirebbe nient'altro che questo: che io ho una ragione in più per desiderare la soddisfazione delle preferenze espresse da qualche sé successivo di quante ne avrei rispetto a un sé che non

sia successivo. Si è detto anche che l'intensità delle preferenze di un sé successivo dovrebbe essere quanto entra nella considerazione relativa alla decisione di soddisfarle. Naturalmente, l'intensità delle preferenze è una delle caratteristiche principali che entrano in gioco quando decidiamo che valga o meno la pena di soddisfare qualche nostro desiderio, ma non è certo l'unica e nemmeno quella più rilevante in ogni circostanza. Noi valutiamo i nostri desideri anche in base ad altri criteri, che sono criteri che sinteticamente possono essere chiamati criteri di virtù e non hanno a che fare semplicemente con la massimizzazione. Piuttosto, tali criteri rappresentano un vincolo alla massimizzazione. Tale vincolo entra come contenuto dei nostri desideri. Mi riferisco ad oggetti quali la lealtà, il desiderio di non manipolare gli altri, di non trarre ingiusto profitto dalle nostre azioni. Tali contenuti hanno a che fare con la nostra descrizione percepita come maggiormente autentica e sono precedenti alla massimizzazione. Ma anche se immaginiamo un agente puramente massimizzante, tale agente avrebbe dovuto pur sempre decidere che la massimizzazione e non altro è ciò che conta come contenuto e oggetto dei propri desideri e tale decisione sarebbe precedente l'altra di massimizzare le proprie preferenze. Sarebbe, in altre parole, una modalità di descrivere l'agente stesso. Nella struttura per così dire trascendentale della propria descrizione è racchiusa una preoccupazione per la coerenza delle serie temporali che si devono riferire a operatori sufficientemente prossimi al sé attuale.

Un agente razionale porrà attenzione al fatto che per quanto è in suo potere, non vengano fatti sorgere dei desideri futuri la cui soddisfazione sarebbe in contrasto con quella che pensa essere la sua autorealizzazione attraverso il tempo. A meno che l'agente non sia animato da una narcisistica e irrealistica volontà di potenza, un agente razionale deve riconoscere che le sue azioni si svolgono e i propri desideri prendono forma in un mondo parzialmente deterministico, nel senso che la possibilità di agire e la soddisfacibilità delle nostre preferenze eccedono la capacità di controllo di chiunque a causa della indefinibilità delle relazioni cooperative coinvolte e dell'incompletezza dell'informazione. Questa notazione del tutto ragionevole può tuttavia generare dei notevoli problemi proprio nella prospettiva della massimizzazione vincolata, che comunque mi pare quella maggiormente adeguata a una concezione narrativa dell'agire. Infatti, poniamo il caso che io sappia che qualcuno agirà o qualcosa avrà luogo con un'alta probabilità nel futuro prossimo, in maniera tale da modificare profondamente il mio carattere e da farmi fare cose che io al momento attuale non farei. L'agente dovrebbe accondiscendere a questa possibilità per il semplice fatto che pensa di non essere in grado di opporvisi? Nella maggior parte dei casi noi immaginiamo che le cose non siano affatto così semplici e non bolliamo come irrazionale un comportamento

anche se si svolge in condizioni molto avverse. Si pensi a quanto è accaduto nei regimi totalitari. La forza di penetrazione e persuasione di tali regimi non è dovuta solo alla loro forza repressiva, ma anche a un fenomeno psicologico che inducono in coloro che sono oppressi: tali regimi appaiono molto stabili. E tuttavia non mancano mai in essi fenomeni variegati e significativi di resistenza. Tale resistenza è motivata dalla convinzione che tali sistemi politici modificano oltre una determinata soglia critica il carattere di coloro che vi resistono. Questi comportamenti smentiscono la linea di minore resistenza, per così dire, nella soddisfazione di preferenze future che non si vorrebbero avere. Vi è un altro esempio che mi pare suggerire che il requisito della coerenza delle preferenze nel corso del tempo rappresenti senz'altro delle stigmate di razionalità impresse all'agente. Supponiamo che per un agente sia coerente avere un gruppo di preferenze *b* nel futuro non analoghe a un altro gruppo di preferenze *a* che ha già soddisfatto nel passato. In un qualche senso, inoltre, l'agente sa che queste preferenze costituiscono parte importante della propria autodescrizione. Tuttavia, decide di non soddisfarle. È sufficiente questo per sostenere che si sta comportando irrazionalmente? Non credo. Infatti, se tali preferenze sono considerate dall'agente come antisociali, potrebbe avere delle ragioni molto valide per non soddisfarle. È innegabile che esistano individui che possiedono tali tendenze ed decidono consapevolmente di non dar loro corso. E magari ognuno di noi si sarà potuto trovare in qualche situazione nella quale avrebbe potuto dar sfogo a un forte impulso anti-sociale senza subirne le conseguenze, eppure non lo ha fatto. Esistono poi anche gli atti di autopunizione che possono essere fatti rientrare in una fattispecie analoga. Tutti questi si presentano come atti non massimizzanti.

Penso che a questo punto sia possibile trarre qualche conclusione almeno provvisoria dalla nostra discussione. Abbiamo visto che la proposta influente di Hare è di assegnare al termine 'io' un significato tanto descrittivo quanto prescrittivo. Si è visto anche che è possibile formalizzare questa posizione esprimendola nella forma di una legge psicologica, la quale afferma che se qualcuno giudica che un sé sia un suo proprio successore psicofisico prossimo, allora è razionale avere come obiettivo la soddisfazione delle sue preferenze in ragione della loro intensità. Si è visto però che esistono ragioni per pensare che questa legge abbia una universalità esclusivamente presunta. Queste ragioni fanno perciò pensare che non sia vera. In effetti, questa legge psicologica, se forse non può essere attribuita ad Hare nella sua forma letterale, gli va ascritta in senso profondo, nella misura in cui realizza una congiunzione tra una posizione fattuale sull'identità personale – la continuità psicofisica dei sé – e una attitudine prescrittiva – quella relativa alla simpatia –. Ciò che si

dovrebbe concludere è che questo legame tra attualità e disposizione alla simpatia non è adeguatamente supportato né in una forma larga né in una forma stretta. Sembrano non esserci ragioni valide per soddisfare delle preferenze sulla sola base del fatto che apparterranno a un successore prossimo; né sembra che ve ne siano per scegliere tra diverse preferenze di un successore sulla sola base del fatto che si dovrebbe scegliere in base alla loro intensità. Appartenenza e massimizzazione possono essere criteri irrilevanti nel decidere la soddisfazione di alcuni desideri di fronte ad altri, proprio in virtù di caratteristiche considerate indispensabili alla propria autodescrizione.

Informazioni sulla rivista / Information on the Journal

Etica & Politica/Ethics & Politics è una rivista filosofica on line, pubblicata in formato elettronico, promossa dal Dipartimento di Filosofia dell'Università degli Studi di Trieste.
L'obiettivo della rivista è di favorire la ricerca e la riflessione, teorica e storica, nell'ambito della filosofia morale e della politica, senza nessuna preclusione culturale.
I contributi dovranno essere sottoposti in una delle seguenti lingue: italiano, francese, inglese, portoghese, spagnolo, tedesco. Tutti gli articoli dovranno essere accompagnati da un abstract in inglese di max. 200 parole.
La redazione sollecita particolarmente contributi interdisciplinari e attenti alle principali tendenze provenienti dal mondo delle pratiche.
La rivista si avvale di un sistema di *referee* anonimo.
Sono previsti due numeri all'anno con cadenza semestrale (giugno e dicembre).
Il **copyright** degli articoli viene lasciato agli autori. A questo proposito, auspichiamo che nei futuri usi degli stessi si menzioni la versione pubblicata su ***Etica & Politica/Ethics & Politics***.

Etica & Politica/Ethics & Politics is a philosophical journal on line, being published only in an electronic format, promoted by the Philosophy Department of the University of Trieste.
The journal aims at promoting research and reflection, both historically and theoretically, in the field of moral and political philosophy, with no cultural preclusion or adhesion to any cultural current.
Contributions should be submitted in one of these languages: Italian, English, French, German, Portuguese, Spanish.
All essays should include an English abstract of max. 200 words.
The editorial staff especially welcomes interdisciplinary contributions with special attention to the main trends of the world of practice.
The journal has an anonymous referee system.
Two issues per year (one every six months: June and December) are expected.
The Author retains all rights, including **copyright**, in the contribution. Of course we will be very glad if, in any future use of the article, the version published on ***Etica & Politica /Ethics & Politics*** is mentioned.

DIREZIONE/EDITOR:
PIERPAOLO MARRONE (Trieste) marrone@units.it

REDAZIONE/EDITORIAL BOARD:
ELVIO BACCARINI (Rijeka) ebaccarini@ffri.hr
ROBERTO FESTA (Trieste) festa@units.it
GIOVANNI GIORGINI (Bologna) giovanni.giorgini@unibo.it
EDOARDO GREBLO (Trieste) grebloe@libero.it
FABIO POLIDORI (Trieste) polidori@units.it

WEBMASTER:
FEDERICO ZIBERNA (Milano) federicoziberna@yahoo.it

COMITATO SCIENTIFICO/ADVISORY BOARD:
A. AGNELLI † (Trieste), J. ALLAN (Otago), G. ALLINEY (Macerata), D. ARDILLI (Modena), K. BALLESTREM (Eichstaet), E. BERTI (Padova), M. BETTETINI (Milano), W. BLOCK (New Orleans), R. CAPORALI (Bologna), G. CATAPANO (Padova), L. COVA (Trieste), S. CREMASCHI (Vercelli), U. CURI (Padova), P. DONATELLI (Roma), P. DONINI (Milano), M. FARAGUNA (Trieste), M. FERRARIS (Torino), L. FLORIDI (Oxford), C. GALLI (Bologna), J. KELEMEN (Budapest), P. KOBAU (Torino), E. LECALDANO (Roma), F. LONGATO (Trieste), E. MANGANARO (Trieste), M. MATULOVIC (Rijeka), N. MISCEVIC (Maribor), R. MORDACCI (Milano), B. DE MORI (Padova), M. PAGANO (Vercelli), A. RIGOBELLO (Roma), P.A. ROVATTI (Trieste), A. RUSSO (Trieste), M. SBISÀ (Trieste), A. SCHIAVELLO (Palermo), F. TRABATTONI (Milano), C. VIGNA (Venezia), S. ZEPPI (Trieste)