METHODS in CS

VI

THE VIRTUAL ISSUE

An Aristotelian Society Publication

METHODS in ETHICS

Guest Editor
Ben Colburn

Henry Sidgwick
J. H. Muirhead
G. F. Stout
S. Alexander
E. E. Constance Jones
G. C. Field
Margaret MacDonald
C. D. Broad
Michael Tanner
Mary Midgley
Onora O'Neill
Robert Stern
Philip Stratton-Lake

Frank Jackson
Brad Hooker
Justin Clarke-Doane
Lorna Finlayson
Robert Cowan
Anna Bergqvist
Roger Crisp
Connie Rosati
Sophie Grace Chappell
Jimmy Lenman

TABLE OF CONTENTS

METHODS IN ETHICS: INTRODUCTION

by Ben Colburn

BEN COLBURN is a senior
lecturer in philosophy at
the University of Glasgow,
where he has worked
since 2010. He earned
his PhD at Cambridge
under the supervision
of Serena Olsaretti and
Hallvard Lillehammer. He
has published on various
topics in ethics, philosophy
of education and political
philosophy, most notably
in his book *Autonomy and
Liberalism* (New York:
Routledge, 2010, pbk
edition 2013).

by Ben Colburn

THIS IS THE THIRD *Virtual Issue* of the Aristotelian Society. Like its predecessors, this volume delves into the archives of the Society's two long-running paper journals: the *Proceedings,* which have recorded the papers delivered at the Society's regular meetings since 1887; and the *Supplementary Volume,* containing the papers given at the annual Joint Session of the Society and the Mind Association since 1918. The Virtual Issues are freely available on the Society's website. The aim is to re-publish pieces in our formidable back-catalogue which merit more attention, either because they are amongst the already-recognised greatest hits of Anglophone philosophy in the last century or so, or because they have come to be somewhat neglected (and therefore deserve to be better known).

This *Virtual Issue*, 'Methods in Ethics', includes a selection of papers from across the Society's fourteen decades, each accompanied by a specially commissioned present-day response. The aim of the volume is to aid reflection on how to go about doing moral and political philosophy, by bringing together papers that bear on their epistemology and methodology. In some cases this involves pieces whose declared focus is on one of those areas, as is the case with Margaret MacDonald's 1941 essay 'The Language of Political Theory' and Mary Midgley's 1974 article 'The Neutrality of the Moral Philosopher'. In other cases I've selected papers whose primary focus is elsewhere, but which include some interesting epistemological or methodological reflections or implications along the way. The contributions from our present-day commentators vary similarly in how far their explicit attention is methodological. Nevertheless, there is throughout a concern (either explicit or implicit) on how to 'be an honest reasoner', to borrow a phrase from the late Dudley Knowles.

The result is an eclectic collection. Some reasons for this are philosophical, and I shall touch on them in the second part of this introduction (wherein I will briefly set out some common themes in the substance of the papers). Others are to do with the way that this volume has been structured. I shall briefly say something about this, since it also illuminates the relationship between the present volume and the history of the Society.

The journals of the Aristotelian Society have always had an unusual character by dint of their recording events – lectures, talks and symposia – rather than simply being venues for publishing philosophy articles. This is especially striking in the earliest issues of the *Proceedings*, wherein we can read extended conversations between members of a comparatively small pool of participants, with the same themes being picked up and examined by speakers in successive meetings. In some cases the *Proceedings* published symposia between several participants, as is the case with the earliest piece in the present collection, a four-way discussion from 1891. These multi-authored pieces quickly disappeared in the *Proceedings*, as the Society adopted the format of a pre-circulated paper delivered by a single author. Still, the symposium form survives in the *Supplementary Volume*, since the main business of the Joint Session is a series of paired papers in which two participants speak on the same topic.

One way of conceiving of the present volume is as something of the same sort: a set of diachronic symposia bringing together interlocutors (sometimes from very different times and places) whose ideas speak to a shared concern. Precisely *how* they speak to each other is rather varied, just as is the case with their synchronic counterparts in the *Supplementary Volumes*. Some contributors have written commentaries on their historical piece, explaining its significance or elucidating it in light of subsequent developments in the subject. Others have written responses to the original arguments, sometimes critical and sometimes supportive. Yet others have treated the earlier piece as inspiration for some related but independent reflections. Different philosophical pieces call for different types of philosophical response.

This diversity of form is matched by diversity of content, and there seems little point in trying to force the particular and varied points made into an overall narrative. In what follows, I briefly introduce each piece in the collection, and then conclude by picking out some shared themes investigated in these symposia which it would be worth highlighting.

The first symposium contains some reflections by Frank Jackson on the reports of a discussion which took place in 1891 between four philosophers: Henry Sidgwick, John Henry Muirhead, George Samuel Alexander, and Frederick Stout. It presents an interesting moment in the history of academic philosophy in Britain. Sidgwick was already a powerful figure: he had made his reputation with *The Methods of Ethics* (first published in 1874), and since 1883 had been Knightbridge Professor of Moral Philosophy at Cambridge. His interlocutors

were all at that stage junior lecturers and fellows, but went on to establish new departments of philosophy at Birmingham (Muirhead) and Manchester (Alexander), and to pioneer the emerging field of academic psychology (Stout, primarily at Oxford and St Andrews).

Sidgwick asked whether, and in what senses, 'ought' claims (about duties) can be reduced to claims about matters of non-normative fact. He distinguished between various different things we might mean by 'reduction': to different purportedly more fundamental theories, for example, or with different attitudes to the persisting status of ethical theory once the reduction has been achieved. Sidgwick thought the reductive project more or less promising depending on these choices. Still, he thought that most reductive projects are problematic because they end up losing something important about the ethical discourse we were trying to understand in the first place: the reductive project 'eviscerates ethical thought of its essential import and interest' (p. 29-30), for example by implying that there is no genuine ethical disagreement because ethical judgements, once properly analysed, cannot conflict.

Muirhead and Stout broadly agreed with Sidgwick. Both made the interesting claim that the ambition to 'reduce' ethical claims to claims about scientific belief is senseless if the motivation is to show how the (potentially problematic) normativity of the former can be made respectable by giving it non-normative foundations. Regardless of whether such a thing is necessary, the reductive project won't provide it, since scientific belief is itself a fundamentally normative matter too.

Alexander disagreed with his colleagues, and sketched a reductive account which he hoped would survive their arguments. His idea was that those objections worked because their targets took the relevant grounding facts to be ones about individual human beings and their judgements or desires, and hence that grounding ethical discourse instead in facts about species flourishing or social functioning (rather than anything directly about individual human beings) would solve the problem. He also responded to Sidgwick's arguments about ethical disagreement by making some nice points about what it means for judgements to conflict, and showing that the phenomenon of genuine ethical disagreement can be preserved by the reductive project he proposed.

Jackson is the new participant in this symposium. He focuses his attention on Sidgwick and Alexander particularly, noting that they pursued their questions without explicit note of the related debates in metaphysics, epistemology and the philosophy of language. One consequence of this is that the opponents of reduction

– Sidgwick, Muirhead and Stout – miss various possibilities which might get past their objections, some of which Jackson sketches here.

One of the striking things about this collection is the way that Henry Sidgwick's reputation waned and waxed over the course of the twentieth century. As Roger Crisp notes in his remarks on Mary Midgley, by the middle of the century he had all but disappeared from view. He has enjoyed something of a renaissance since – Derek Parfit calls *Methods of Ethics* the 'best book on ethics ever written', and he isn't alone in his praise – but Sidgwick's reputation has never been higher than in the decades after his death in 1900, when he was a titanic presence. We can see this in Constance Jones's 1917 essay 'Practical Dualism'. Jones, mistress of Girton College Cambridge and an influential logician and philosopher of language, here produced an almost entirely exegetical essay, defending and elucidating Sidgwick's view that rational benevolence and rational self-love are the two core principles of morality, and that (perhaps contrary to appearances) they are compatible, with the former implying or including the latter. The most interesting elements of her piece, from our point of view, have to do with her remarks on the relationship between utilitarianism and the rules of common-sense morality.

Hooker, in his response, articulates a general worry about the grounding of utilitarian theories. Some utilitarians have a tendency to pull the rabbit out of the hat: the scientistic trappings with which such theories are often defended have a tendency to make it appear that the (substantive) normative conclusions can be derived without any substantive normative assumptions being made. He thinks that Jones was guilty of this. That raises an interesting point for us, of course, since the question of how to capture the normative force of (utilitarian) ethical theory was one of Sidgwick's central concerns. In this respect, perhaps, Jones is more similar to earlier and less sophisticated utilitarians like John Stuart Mill, rather than Sidgwick himself.

Hooker notes Jones's methodological remarks, but he modestly refrains from elaborating on the interesting parallels that might be drawn with his own important work on rule consequentialism (e.g. Hooker 2000, and see also Brandt 1963). The type of distinction Jones defended was one which separated a realm of critical philosophy (in which we affirm utilitarianism) from a realm of everyday action (in which we may habitually follow common sense morality). But it is worth noting that we can construe her remarks (and Sidgwick's) as indicating something different: not a distinction between higher- and lower-order domains, but a certain internal com-

plexity to the first-order moral theory itself. As Hooker notes, the crucial move is to build on Sidgwick's insight that we can 'evaluate codes of rules by their utility', and find grounds for taking that approach in common-sense morality.

In his 1931 essay, Guy Cromwell Field – a one-time spy and prisoner of war, but for most of his career Professor of Philosophy at Bristol – contrasted practices of definition in ethics with those in other fields. His piece pursues some interesting parallels between ethical definition, stipulative definitions in geometry, and functional definitions in biology. Field's central concern was to argue against attempting to define terms too sharply to begin with, in favour of a procedure which (as Justin Clarke-Doane comments in his reply) looks much like John Rawls's later method of reflective equilibrium (see e.g. Rawls 1971).

Clarke-Doane takes up the challenge and plays further with the mathematical comparison. He considers whether the formal devices developed by philosophers of mathematics, or their commentary on the subject's epistemological or metaphysical status, might be harnessed to help ethicists deal with the parallel worries directed at them. His ultimate conclusion is pessimistic, but the reasons for failure illuminate some important features of ethical discourse that might be harder to see without the comparison with mathematics.

Margaret MacDonald was a student of Susan Stebbing, G.E. Moore and Ludwig Wittgenstein, but she is most closely associated with the last of these. Her 1940 piece on the language of political theory is a Wittgensteinian perspective on how political philosophers typically talk, and how political philosophy relates to practice and history. MacDonald's piece is full of insight and clarity. One of her central concerns was to identify and diagnose a stultifying tendency in much political theory to resort to excessive idealization or metaphysics. This, she argued, has a tendency to kick real political questions into the long grass, and to make us too inclined to trust the political status quo (and its government). It is interesting to see MacDonald independently developing some of the lines of analysis which have subsequently been more famously associated with the Frankfurt School of critical theory, at that point in its infancy on the continent.

Lorna Finlayson, in her response, reflects on some further philosophical lessons we might draw from MacDonald's piece. One has to do not with the content of her essay but with the stance she occupies, of the outsider: Finlayson remarks that there is much to be gained, as well as lost, by refusing to cast oneself as an expert with-

in a specialised discourse. She also discusses the extent to which MacDonald might be regarded as a very early exponent of 'realism' in political philosophy, which is to say the anti-utopian view that the political philosopher ought to spend less time discussing idealized 'well-ordered societies', and more time analysing real political conflicts and relations of power (see, for example, Williams 2005, Geuss 2009, and Rossi & Sleat 2014). The parallels are interesting, since realism is usually cast as a reaction to the influence of Rawls, whose main works postdate MacDonald's untimely death in 1956.

Charles Dunbar Broad was, after some time working elsewhere, a successor of Sidgwick as Knightbridge Professor of Moral Philosophy at Cambridge. His 1944 essay explores ethical theories which explain our knowledge of morals by appealing to a 'moral sense'. Such theories are associated primarily with the early modern Sentimentalists, but have enjoyed various revivals during the twentieth century. With characteristic precision and attention to detail, Broad distinguished between various different ways that one might construe this epistemological theory of moral judgements. He also reflected on how it relates to the analytic question of how we are to understand the content of those judgements.

Broad was in many ways ahead of his time. Peter Smith, in his Routledge Encyclopedia biography, remarks that 'where more recent writers have overlooked his writings, such as those on moral-sense theories, it has often been at the cost of labouring to rediscover distinctions already elegantly made by Broad.' Robert Cowan's piece here helps to bridge the gap; he goes toe-to-toe with various of Broad's arguments, and then explores what subsequent philosophical developments (for example in the philosophy of perception) reveal about the prospects for the sort of neo-Sentimentalism which Broad considered.

Michael Tanner is mostly known for his work in aesthetics, literature and musical criticism. His 1964 piece draws on these broad interests to make some interestingly prescient points about ethics. The essay is a reaction against the contemporary focus on a very narrow range of purportedly fundamental ethical terms: 'good', 'right', 'ought' and 'duty'. Tanner complained that an exclusive focus on these had resulted in 'moral philosophy [that is] parochial, partial, boring and irrelevant to practice', and that moral philosophers ought instead to grapple with the whole range of concepts which we use in our ethical lives, including those whose rich content is really only brought out by considering very detained and concrete examples.

As Anna Bergqvist points out in her response, Tan-

ner's piece anticipated much subsequent work on the relation between what are now called 'thick' and 'thin' moral concepts. She shows that Tanner held a sophisticated version of the thought that the morally relevant facts cannot be accessed save through some individual perspective. This commitment – inspired by Tanner's work in aesthetics – has important consequences for the familiar distinctions between fact and value, and between the subjective and the objective.

Mary Midgley, writing in 1974, picked up a theme also present in the contributions by MacDonald, Finlayson, and Tanner. The question is: should the moral philosopher be neutral (in some sense or other) concerning the moral views that she examines? Midgley – with Charles Stevenson as her stalking horse – thought that a constraint of neutrality was unattractive and confused; it would have the effect of distracting our attention, and of distorting our first-order moral theory. So, those who advocate neutrality on the basis of a desire to be fair and detached should think again, and adopt instead an ideal of impartiality, which is consistent with openly recognising one's partial, partisan viewpoint.

Roger Crisp draws out various points of interest in Midgley's paper, but notes that she is vulnerable to a sort of self-defeat: the position of impartiality that she advocates will often lead to an even greater distortion of one's moral view than a mistaken adherence to neutrality. He suggests that we might get further resources for a *defensible* neutrality, immune to Midgley's arguments, by noting that some methodological questions about ethics (what we start with, what we treat as central, and so on) might be justified not morally but epistemically, and that this might provide a kind of neutral space within which discussion should proceed.

Several of the essays described so far are concerned (sometimes implicitly) with a central dilemma of metaethics. Moral realism is often taken to be unattractively costly, for example because it is metaphysically extravagant, or committed to an implausible moral epistemology. Is it possible, though, to avoid these costs without being committed to something equally unattractive, namely a sort of relativism or subjectivism which (to repeat Sidgwick's words) 'eviscerates ethical thought of its essential import and interest'?

Constructivism is a position in metaethics which seeks to navigate between these dangers. We analyse our ethical claims as being reports of what would be the outcome of some (usually idealized) process of deliberation or agreement. This avoids the unpleasant metaphysical and epistemological price-tag of a more traditional Platonistic realism, while at the same time it incorporates

sufficient critical bite to avoid collapsing into relativism. That, at any rate, is the ambition. In her 1988 essay, Onora O'Neill scrutinizes John Rawls's famous attempt to defend an account of this sort, and argues that it fails to live up to the ambition. She then goes on to sketch a version of constructivism which, in her view, solves these problems, by returning to a more authentically Kantian position.

Connie Rosati's response suggests that O'Neill's view doesn't escape so lightly, especially once we attend closely to the notions of idealization involved in specifying the constructive procedure. She argues that, on closer inspection, Rawls's position might after all have more going for it than O'Neill credited.

Robert Stern's 1992 essay addresses one of the central topics of this volume, the relation between moral philosophy and other branches of the discipline. Stern's target is Rawls's view (e.g. in Rawls 1974) that moral philosophy should be treated as quite independent of metaphysics, in the sense that the latter cannot settle questions within the former. One of Rawls's principal tactics was to identify purportedly metaphysical claims with serious moral theoretic weight, and to argue that the claims in question are in fact disguised claims *within* moral theory, and not genuinely metaphysical at all. Stern begs to differ, offering reasons to reject this central claim of Rawls's, and in the process making some shrewd obervations about the method of reflective equilibrium on which Rawls's moral and political philosophy rests.

Sophie Grace Chappell picks up the theme, but argues that developments in the philosophy of mind are even more important for our ethical theory. She gives a number of examples of cases where this might happen, culminating with a discussion of the role of thick concepts in our ethical lives. It is interesting to compare Chappell's reasoning here with Michael Tanner's earlier arguments along similar lines.

The concluding pieces in the symposium speak to a recent revival of interest in intuitionism, which is to say the position in moral epistemology which says that ethical propositions are in some sense self-evident and can be known non-inferentially. This is not unrelated to the topic of Broad's paper earlier in the symposium, but – as indeed is demonstrated by Cowan's piece in reply to Broad's – modern moral epistemologists take advantage of various developments in philosophy from outside ethics to make clearer what might be needed for this sort of account of ethical knowledge to work.

Philip Stratton-Lake's 1999 essay, and James Lenman's response, speak to this debate. Stratton-Lake

considers arguments against intuitionism that hinge on pointing out that it is a position committed to some form of externalism, which is to say (roughly) a variant of the idea that there is no essential connection between moral judgement and motivation. Stratton-Lake agrees that intuitionism is so committed, but suggests that on reflection this need not be something which weakens intuitionism. Lenman begs to differ. He notes that there might be something to be said for intuitionism or externalism taken separately, but that the conjunction of the two seems to commit us to some unpleasant (indeed unintelligible) views about what normativity really amounts to.

These essays represent a century and a quarter of moral philosophy. They are very diverse in their attention, but it seems to me that there are three central questions which most of them touch upon in some form or another. One is to do with the epistemology of ethics: how, given the distinctive features of ethical discourse, might we give an account of moral knowledge? (e.g. Sidgwick, Field, Clarke-Doane, Broad, and Cowan). Another is to do with metaphysics: can ethical thought be accommodated within a broadly naturalistic framework, and if so, how? (e.g. Alexander, Jackson, Field, O'Neill and Rosati). A third, related to both the first and second, is: is it possible for ethical judgements to have real critical force, while also being appropriately connected to motivation? (e.g. Sidgwick, Jones, Chappell, Stratton-Lake and Lenman). Many of the answers offered here are familiar from the contemporary philosophical scene; but it is interesting to see them appear earlier, and from different authors.

Also present throughout these essays, but less part of their explicit focus, are some more synoptic methodological concerns. I want to conclude by drawing out two of these.

One has to do with the relationship between moral and political philosophy and the rest of the discipline. Several of our authors suggest that we can shed light on ethical questions by drawing inspiration from, or deploying conclusions from, other areas of the subject (e.g. Field, Clarke-Doane, Tanner, Stern, and Chappell). Moreover, a point made by various of our twentieth-first century symposiasts is that their earlier interlocutors might have done better if they had done more to draw on the broader philosophical context for their thinking about ethics (e.g. Jackson, Hooker, and Cowan). In general it seems that moral philosophers have become somewhat happier drawing on the tools of theoretical analytic philosophy to shed light on ethical questions. This is encouraging. It also undermines the view that

philosophy as a whole is suffering from its practitioners' increasing specialization, by comparison with an earlier golden age of philosophical generalists. Many of the earliest writers here were remarkable polymaths in their interests (especially Sidgwick, Stout, Alexander and Jones), but (in the pieces represented here, anyway) they seemed to treat moral philosophy as autonomous, and drew little from their expertise in metaphysics, the philosophy of mind, and logic.

The other has to do with ideals of neutrality in moral and political philosophy. This is an explicit focus of the essays by MacDonald, Finlayson, Midgley and Crisp. But it is also lurking under the surface in some of the others, for example in Tanner (as Bergqvist makes explicit in her commentary). There is plainly something attractive about the idea that philosophers should operate under *some sort* of norm of neutrality when writing about ethics or politics. But there is a danger that this commitment – especially if left implicit – will end up serving some covert ideological purpose, by adversely affecting our judgements of salience, of the burdens of proof, and of what counts as an open question. The effect of this might just be to make our philosophy more limited and parochial than we thought (as Crisp observes of Midgley): that would be regrettable, but not catastrophic. But it might also leave us vulnerable to cooption by some political forces which we would otherwise be inclined to repudiate, as MacDonald and Finlayson worry. How exactly to avert these dangers is not an easy question to answer; but I think the present collection gives reason to think that grappling honestly with questions about method in ethics is a good start.

REFERENCES

Brandt. R. (1963), 'Towards a Credible Form of Utilitarianism', in H.-N. Castañeda and G. Nakhnikian (eds.) *Morality and the Language of Conduct*. Detroit: Wayne State University Press, pp. 107–43.

Geuss, R. (2009), *Philosophy and Real Politics*. Princeton: Princeton University Press.

Hooker, B. (2000), *Ideal Code, Real World*. New York: Oxford University Press.

Rawls, J. (1971), *A Theory of Justice*. Cambridge MA: Belknap Press.

———— (1974), 'The Independence of Moral Theory', *Proceedings and Addresses of the American Philosophical Association* 48: pp. 5-22.

Rossi, E. & Sleat, M. (2014), 'Realism in normative political theory', *Philosophy Compass* 9/10: 689-701.

Smith. P. (2015) 'Charles Dunbar Broad' in E.J. Craig ed. *Routledge Encyclopedia of Philosophy Online*. (New York: Routledge). https://www.rep.routledge.com/articles/broad-charlie-dunbar-1887-1971. Accessed 1 September 2015.

Williams, B. (2005), *In the Beginning was the Deed: Realism and Moralism in Political Argument*. Princeton: Princeton University Press.

IS THE DISTINCTION BETWEEN "IS" AND "OUGHT" ULTIMATE AND IRREDUCIBLE?

A symposium featuring
Henry Sidgwick
J. H. Muirhead
G. F. Stout
S. Alexander

HENRY SIDGWICK (1838 – 1900) was born in Skipton. He was educated at Cambridge, where he eventually spent his entire career, first as a lecturer in classics and a fellow of Trinity (till he resigned in 1869 because he could no longer subscribe to the Thirty-Nine Articles of the Church of England), then as one of the first members of the Faculty of Moral Sciences, and finally from 1883 as the Knightbridge Professor of Moral Philosophy. He was a towering presence in Victorian and early twentieth century Anglophone philosophy and (after a period of comparative neglect in the mid-twentieth century) still exerts a wide influence on many working in moral philosophy. He worked widely on moral philosophy, political philosophy, economics, politics, literary theory and educational theory, and was a prominent voice for reform (especially concerning the place of women in universities). His most famous work is The Methods of Ethics, first published in 1874. Derek Parfit calls it 'the best book on ethics ever written'.

JOHN HENRY MUIRHEAD (1855 – 1940) was born and educated in Glasgow. In 1875 he moved to study further in Oxford, and then taught at Glasgow and Royal Holloway before settling at the new University in Birmingham where in 1900 he became the first Professor of Philosophy and Political Economy. He retired in 1921, but remained an active writer and lecturer in the UK and the US. Muirhead wrote in ethics, ancient philosophy and politics, but is probably best remembered for masterminding the Muirhead Library of Philosophy, a distinguished series in which many seminal works were first published.

GEORGE FREDERICK STOUT (1860 – 1944) was born in South Shields and studied at Cambridge. He worked there and at Aberdeen before becoming successively Reader in Mental Philosophy at Oxford (1898) and Professor of Logic and Metaphysics at St Andrews (1903). He stayed at St Andrews till his retirement in 1936. Stout's work was in the intersection of philosophy and psychology; he was editor of the distinguished journal Mind from 1891 to 1920, and his best known works are in psychology: the Analytical Psychology of 1896 and his Manual of Psychology of 1898.

SAMUEL ALEXANDER (1859 – 1938) was born in Sydney. He started studying at the University of Melbourne, but left in 1877 to take up a scholarship at Oxford. On graduating he became a fellow of Lincoln College in 1882, the first

Jew to obtain a fellowship in either Oxford or Cambridge. In 1893 he was elected Professor of Philosophy at Owens College, which later became the Victoria University of Manchester, and reigned there till his retirement in 1924. He published variously in metaphysics and aesthetics, and also spent much time seeking to advance the position of women in higher education.

A symposium featuring
Henry Sidgwick
J. H. Muirhead
G. F. Stout
S. Alexander

I. HENRY SIDGWICK

1. I SHALL ASSUME AT THE OUTSET of this necessarily short paper, that we are generally agreed as to the objects of thought to which the predicates "is" and "ought to be" are respectively appropriate; though I shall have occasion in the course of the paper to notice certain variations of opinion on both points. I shall also take "what ought to be" to include what is commonly judged to be "good," so far as attainable by human action, as well as what is commonly judged to be "right," or "the duty" of any human being. Of course "good" and "evil" as commonly used are wider and less stringent terms than "right" and "wrong"; since (1) the former are applicable to results out of the reach of human attainment – an abundant harvest next autumn or influenza in the winter – also (2) "goods" may be incompatible, to attain a greater we may have to sacrifice a less. But even when unattainable, or not preferable in the circumstances, what is judged to be "good" would appear to have the same quality as the term imports within the range of its practical application; "good" is the kind of thing that we "ought" to seek to produce or maintain *pro tanto* and so far as it is in our power.

For simplicity, I shall mean by "good" in this discussion, "ultimate good on the whole"; and to avoid complicating the discussion, I shall assume that what is good on the whole for any individual agent is also good on the whole for human society, the world of living things, or the cosmos, whichever we take to be the larger whole of which the individual is a part, and which is conceived to have an ultimate good capable of being increased or diminished, promoted or retarded, by human action. That is, I shall assume that "what ought to be" is the same from the point of view of self-interest and from that of duty. The notion of "right " or "duty" is, however, more familiar in ethical discussion to the common moral consciousness of modern men – to which I shall refer as common sense – than the notion of "ultimate good." But I shall assume it to be admitted by common sense that, from the point of view of complete knowledge, the performance of a duty or a right act must be conceived to be either a part of ultimate good or a means to it.

Taking then the notion of Duty or Right act – I may assume it to be a continually recurrent element in the thought of an ordinary well-behaved person about his own life and that of others. In the thoughts of such men about duties, taken together and compared, there is doubtless more conflict and disagreement than in their thought about facts; but agreement much preponderates. Apart from such conflict, there is a recognised variation of duties from man to man, but it is commonly assumed that this variation is on Rational grounds, so that the duties of A, truly conceived, form one rationally coherent system with the duties of B. Such a system we may call a "world of human duty," of which each man conceives the duties he assigns to himself and his immediate neighbours to be a part indefinitely better known to him than the rest; but he conceives the whole world of duty to be a subject of human knowledge, no less than the world of fact, though the former is lamentably divergent from the latter, in consequence of the general failure of men, in a greater or less degree, to do their duty. The divergence is equally palpable if we consider the "good" results that might be brought about by the performance of duty, as compared with what actually takes place. From either point of view we judge that "what ought to be" to a great extent "is not," and we commonly conceive that its character as "what ought to be" is entirely independent of whether it comes into actual being, or not.

2. The question then is raised, whether this distinction between what is and what ought to be is ultimate and irreducible? I think it rash to affirm irreducibility, but I am certainly not satisfied with any proposed reduction proceeding on the lines of scientific thought on which such reduction is commonly attempted; i.e., I do not think the desired result can be attained by considering moral judgments from a psychological or sociological point of view, as elements in the conscious life of individuals or communities or races. No doubt moral judgments and their accompanying sentiments are a department of psychical fact, and we may analyse and classify them as such and investigate their causes, just as we should do in the case of any other psychical fact; but as long as they are regarded from this point of view, it seems impossible to explain or justify the fundamental assumption on which they all proceed, that some such judgments are true and others false, and that when any two such judgments conflict, one or both must be erroneous. One fact cannot be inconsistent with another fact; accordingly, regarded from a psychological or sociological point of view, A's judgment, e.g., that all gambling is wrong, does not conflict with B's judgment that some gambling is right; the question, which is true, does not arise and would have no meaning. The reduction, therefore, of duty to fact, on this line of thought if strictly pursued, eviscerates ethical

thought of its essential import and interest.

It may be replied, perhaps, that in this argument I have not taken into account of the notions of life and development, and their place in psychology and sociology, that possessing these notions, science in this department does not merely ascertain resemblances and general laws of co-existence and change, but in so doing brings out the notion of an end to which psychical and social changes are related as means, and in relation to which alone they are really intelligible; and that this end supplies the requisite reduction of "what ought to be" to what is. For in this end – variously conceived as vital or social "health" or "equilibrium" or "life measured in breath as well as length," we have (it is thought) a criterion of truth and error in moral judgments; if the acts they approve are conducive to this end, they may be counted true or normal, if not, false or abnormal.

To this I answer that End as a biological or sociological notion may, no doubt, be held convertible for practical purposes with ethical end; but that this can only be by an ethical judgment affirming the coincidence of the two: the two notions remain essentially distinct, though when affirmed to be coincident they are doubtless liable to be confused. From the mere knowledge that a certain result is what will be or preponderantly tends to be, it is impossible to infer that it ought to be; so far as it is inevitable, I obviously can have no duty with regard to it; so far as its coming may be promoted or retarded, it is my duty to promote it if I judge it good in comparison with that for which it would be substituted, and to retard it if I judge it to be comparatively bad. Perhaps I may suggest as a reason why this is often not clearly recognised, that in the terms such as "social welfare" or "social health," used to denote the sociological end, the ethical notion is surreptitiously introduced; they are states which have been implicitly judged to be good.

3. When I turn from the point of view of Science to that of Philosophy or Epistemology, before answering the question whether the difference between what is and what ought to be is irreducible, I require to know exactly what is meant by "reduction." Is the difference between two things reduced by merely discovering previously unknown resemblances between them? E.g., we may compare the circle and the parabola without knowing that they are both sections of the cone; should we say that the difference between them ascertained by this comparison is reduced by discovering their common relation to the cone? If so, I think it must be admitted that this kind of "reduction" takes place when we contemplate the difference between "what is" and "what ought to be" from a philosophical or epistemological point of view. For from

this point of view we regard the world of duty and the world of fact alike as objects of thought, and – real or supposed – knowledge, and discover similar relations of thought in both, relations of universal to particular and individual notions and judgments, of inductive to deductive method, etc.; whatever differences may appear between the two from this point of view are of a subordinate kind, and not greater than the differences between different departments of fact regarded as objects of thought and scientific method. True, if we adhere to common sense, the fundamental difference remains that the distinction between "truth" and "error" in our thought about what is, is held to depend essentially on the correspondence, or want of correspondence, between Thought and Fact; whereas, in the case of "what ought to be," truth and error cannot be conceived to depend on any similar relation except on a certain theological view of duty, which I will presently notice. Still, even this difference is at least reduced if we take the philosophical point of view; because, from this point of view, the supposed correspondence between Thought and what is not Thought is no longer so simple and intelligible as it seems to common sense; it must be recognised as a difficult problem, whatever solution of it we may ultimately accept. It must be recognised that, even in the case of our thought about what is, though error may lie in want of correspondence between Thought and Fact, it can only be shown and ascertained by showing inconsistency between Thought and Thought, i.e., precisely as error is shown in the case of our thought about what ought to be.

Perhaps, too, the difference between "what is" and "what ought to be" may be reasonably held to be relatively reduced, when we contemplate, along with both, various forms of "what might be" or "might have been," as objects of more or less coherent thought for scientific or artistic purposes.

4. Finally, I must notice another method of "reduction," at first sight plausible and more near to common sense than the philosophical. This proceeds on the theological assumption that the true rules of duty are Divine commands – whether made known by external revelation or through the conscience of the individual. Such commands, it is said, may be imperfectly known to any particular moral agent, either without his own fault – in which case their non-fulfilment will be pardoned – or through wilful neglect of known duty in the past, which has had the effect of impairing his moral insight; but in any case such commands have been uttered, and must be regarded as a part of universal fact. I think, however, that this reduction fails when we work it out. Firstly, we cannot define a Divine command – like a human com-

mand – as a wish combined with a threat – since we cannot attribute to God an ungratified wish. Shall we, then, conceive it simply as a threat? This would clearly offend common sense, which conceives God as not merely an "Omnipotent Ruler," but also a Righteous Ruler, commanding in accordance with a rule of Right.

But thus the difference we are considering emerges again in the form of a distinction between the Rule of Right in the Divine Mind and the Divine Power as manifested in the world of fact; and, emerging, it brings with it the formidable problem of the existence of evil, since we inevitably ask why God's power does not cause the complete realisation of ideal Right and Good. This question has received various answers, but it is hard to find an answer which does not maintain unreduced the difference between "what is" and "what ought to be."

II. J. H. MUIRHEAD

IN WHAT I AM GOING TO WRITE on this question, I do not propose to criticise Professor Sidgwick's paper. I agree with his contention that any attempt to reduce the "ought" to the "is" "on the lines of scientific thought" can only end in "eviscerating ethical thought of its essential import and interest."

In this paper I propose to take up the question where Professor Sidgwick leaves it in Section II of his paper, and to ask, if we reject the naive mode of identification there criticised, are we then to regard the "is" and the "ought" the assertion of fact, and the categorical imperative as wholly independent of one another? Or if not, what are we to conceive their relations to one another to be? In trying to answer this question, I may begin by pointing out that ethical writers often speak as though these categories stood to each other in merely a negative relation. Thus in an article in a recent number of the *International Journal of Ethics* (vol. i, No. 1), which I have before me, I find the "ought" of morality sharply contrasted with the "is" of science. "The maxim," it is maintained, "'Do unto others what you would that they should do to you,' does not indicate of itself what happens or ever has happened, or ever will happen – it is a rule prescribing what should happen. It is not gathered from experience, or founded on experience; it is a demand of the mind." This contrast is further emphasised by argument to show:

(1) That ideas of what ought to be are not verifiable. Ideas as to what exists or what happens "are all actually or conceivably verifiable. … But how can we verify ideas, not as to what happens or exists, but as to what ought to happen? . . . We have to believe in them, if we

believe in them at all, not because they have the fact on their side, but because of their own intrinsic attractions and authority."

(2) Ideas of what ought to be require no explanation. "All that happens, that begins to be, requires explanation. All matter-of-fact laws, like gravitation, chemical affinity, and the like .... may find their *raison d'être* outside themselves: may exist, for example, ultimately for moral ends, but the moral laws . . . do not exist for ends beyond themselves, but to dominate all other ends." Here we have in its sharpest form the distinction against which the suggestion in the question at the head of this paper – that the "ought" and the "is" may ultimately be reducible to one another – is presumably directed. Although the distinction itself, as drawn out by the writer, from whom I have quoted, is conceived of in a wholly superficial manner, I agree with him in holding that there is a contrast and distinction, and that it is as wide as the distinction between knowledge and practice, thought and will, the real and the ideal. But this cannot be taken to mean, as the above passages seem to imply, that the moral imperative is intelligible, apart from the facts of the existing moral order. There is, of course, a sense in which it is true, as is contended, that the ought is a "demand of the mind." It is only in so far as the facts of the moral order reflect themselves in the mind and conscience of the individual that they can appeal to him as a moral being. But while it is thus true that duty is prescribed by the *mind*, it is more important for our present purpose to observe that it is prescribed by the *facts* as well, and that it never could be prescribed by the mind unless it were prescribed by the facts. Hence the prime duty of an intelligent analysis and apprehension by the individual of the particular circumstances in which he is called upon to act. No one can perform this analysis for him. Teachers and preachers may indeed provide general maxims such as "Thou shalt not steal," "Thou shalt not kill," "Tbou shalt love thy neighbour as thyself," but, as has been well said, these are merely "tools of analysis." The analysis has yet to be made, and it can only be made by the individual who has in view the details of the situation in all its particularity as it presents itself to him. What blinds us to the necessity for this minute analysis of the facts, apart from our general moral obtuseness, is that in ordinary life our duties are comparatively plain. A great deal is said in books on Ethics about conflicts of motives, but in common life these conflicts are the exception, not the rule. When such a conflict actually takes place, and we become engaged in what is called a moral struggle, we may easily observe this process of analysis going on. Take the case (to borrow Professor Sidgwick's illustration) of a gambler, whose conscience is roused as he handles his last 10,000 francs: shall he stake it, or

shall he not? Forthwith the analysis is set on. On the one hand there is the calculation of the chances of his winning, all that he might do with the money if he had it again, there is the pleasure of it, it is the plucky thing to do, and so on. On the other hand there is the chance of losing, the ruin and disgrace, the misery of his family, etc. – the whole process giving, let us say, as the final result, the duty to keep his money in his pocket. When this point is reached, the moral imperative ought not to be conceived of as something superimposed upon the facts, or setting itself in opposition to them. The "ought" is not something superimposed upon the "is"; it comes out of it; it may be said to have been given in it, and in this sense may be said to be already a kind of new "is"; it is the "is" in the making, or the "is to be"; it is the is of the fact conceived of as living and moving as opposed to the ordinary "is" which has been defined in this same connection as "the 'is' of the fact at rest."[1]

But not only may we say that the ought rises out of and falls back into the is in the manner just explained; we may go a step further and say that, inasmuch as the real nature of anything is that which it has in it to become, rather than that which it already is, the essence of human relations (the *facts* of the moral world) and the society which is built up out of them is to be looked for, not in man's actual achievements in these respects, but in the end or ideal towards which he is progressing. It is this which gives them their form; it is their final cause, and as such determines them to be what they are. As Aristotle said that "in the order of nature, the state is prior to the household or the individual, for the whole must be prior to the parts"; so we may perhaps say that that towards which human society is evolving, that which it "ought to be" is prior in the order of nature to that which already is and constituting its true essence or "is-ness." And this dependence of the "is" upon the "is-to-be" is still more obvious if for the objective we substitute the subjective point of view, and look at the matter from the side of the individual will, and the ideal or "ought-to-be" which it sets before itself. For just as it is true that in the world of idea the system of truth, which we call science, only exists in virtue of the courageous and persevering effort of individuals to comprehend and extend it, and without this would be an unintelligible jargon extending over innumerable books and burdening the shelves of innumerable college libraries, so in the world of will it may be shown that the actual system of moral relations, as we know them, is only sustained by the loyalty of individuals to the duties, which from moment to moment present themselves. But as this loyalty is not, strictly speaking, to anything that "is," but to an "ought to be," it is perhaps not an exaggeration to say that the fabric of our family, university, estate, and cosmopolitan institutions, which,

1.   Dewey, *Journal of Ethics*, i, 2.

looked at from the outside, seems to rest on so solid a foundation, looked at from within, does not, strictly speaking, rest upon anything that is real at all, but upon the ideal structure of the "ought."

The point I have been endeavouring to put may finally be illustrated from the field of science itself. There is no one who resents more keenly the imputation of being an idealist than the man of science. Yet it only requires a moment's consideration to see that, while conceiving himself emancipated from any allegiance, save allegiance to fact, he is really the slave of an ideal, or of something which is not as yet, but which he recognises ought to be. For while it is undoubtedly true that science deals with actual facts, it does so in the interest of an ideal order, not yet actually apprehended, but believed by all workers in this field to be in the long run apprehensible. In other words, science is as much concerned with what ought to be as with what actually is. The conception of what ought to be is undoubtedly based upon what is. Like the ideal of morality, the ideal of science is dependent for its form and features upon experience of the real. But on the other hand, the actual data are, as it were, held in solution to be cast into a new form at the bidding of the ideal. Thus the astronomer, who marks a hitherto unobserved eccentricity in the orbit of a planet, knows that "there ought to be" some cause for it, some neighbouring body perhaps hitherto unknown. This ought to be is, of course, in its leading features, determined by what is: the body must be in such and such a direction, must be visible if at all at such and such a moment. But on the other hand it reacts upon the observer's view of the facts already known, and in the interval between the suggestion of the existence of the new planet and its actual discovery, his mind is engaged in reconstituting the existing data, so as to form, in conjunction with the new cause which his ideal demands, a whole which will be consistent with itself.

There is, of course, an important difference between the ideal or the "ought to be" of science, and the ideal or "ought to be" of morality. In the former case we conceive of the relation to be discovered as already existing in nature. It exists objectively, it does not yet exist subjectively, save in the form of a hypothesis. In morality the order is reversed: the ideal is not yet anywhere wholly realised as an objective fact; on the other hand there is a sense in which it may be said to be subjectively realised: the good will in the individual already contains potentially the new social order for which it works. He does not need to wait to see it realised in the objective world of fact. The Kingdom of Heaven is in a literal sense within him.

Nevertheless, the parallel is sufficiently close to illustrate the relation of mutual dependence upon one another, which I have tried to establish between the "is" and the "ought to be." The "ought" is dependent for its form upon the "is"; the "is" is from moment to moment sustained and reconstituted by the action upon it of the "ought."

III. G. F. STOUT

PROFESSOR SIDGWICK SAYS: "It must be recognised that even in the case of our thought about what is, though error may lie in want of correspondence between Thought and Fact, it can only be shown and ascertained by showing inconsistency between Thought and Thought, i.e., precisely as error is shown in the case of our thought about what ought to be." Following the hint here given, I shall begin by asking, What is meant by the word ought, when it is said that we ought to *believe* a thing? In other words I shall treat first of the theoretical ideal, and *Truth* before proceeding to deal with the practical ideal, *Right*, or *Good*. Now, from the theoretical point of view there are two ways, and only two ways, in which a man may be brought to see that he ought not to hold this or that belief. In the first place, he may find that it is inconsistent with his other beliefs so as to violate the systematic unity of his pre-formed view of the world. In the second place, though there may not at the outset be any difficulty in fitting the belief into what we may call his intellectual preformation, nevertheless he may ultimately be constrained to reject it because in the course of further experience he acquires new beliefs with which it is irreconcilable. Here the test of truth is not mere self-consistency, but self-consistency conjoined with extension of his experience. We ought to endeavour to disbelieve whatever is irreconcilable with the systematic unity of a progressive experience, or to vary the emphasis, whatever is irreconcilable with the progressive enlargement of a unified experience. Of course, in applying this principle to convince a man that he is wrong in holding a certain view, we can only appeal to the circle of ideas which he has already acquired, or to such an enlargement of it as he may obtain by following our directions or listening to our testimony, or to our instruction.

Mutatis mutandis, the acquisition of truth and the elimination of error in our own case is possible only by a similar process. It would seem then that in the sphere of theoretical activity the distinction between what is and what ought to be is not irreducible. To say that this or that ought not to be believed, is to say that it will in the long run be irreconcilable with the systematic unity of an ever-growing experience. To say that it ought to be

believed is to say that its contradictory will be similarly incapable of forming an integral part of the ever-widening circle of ideas. But this reduction is based upon an assumption. It presupposes that men either necessarily pursue truth or that they *ought* to do so. Now in fact it is certain that men do not universally and necessarily pursue truth. "Where ignorance is bliss, 'tis folly to be wise," is a saying which expresses a point of view adopted more or less consciously by a large part of mankind. But if we say that truth *ought* to be pursued the question arises, how is this *ought* reducible to an *is*? If a satisfactory answer is to be found, it obviously can only be found by reducing the practical *ought*, as distinguished from the theoretical, to an *is*. For the primary obligation within the sphere of theoretical activity – the obligation to seek truth – is a moral obligation.

Turning from theory to practice, we easily find a formula expressing what a man ought to do corresponding to the formula expressing what a man ought to believe. A man ought to endeavour not to aim at such ends as will be found ultimately inconsistent with the systematic unity and increasing range and depth of practical interests. The meaning and justification of this formula will be best brought out by showing its application in particular cases.

Suppose that I have to do with a burglar who takes a pride in his profession and pursues it as his chief aim in life. If I wish him to be convinced that he ought not to get drunk frequently and inconsiderately, I may do so by pointing out that it leads to oversights and bungling, which spoil his career as a housebreaker. If he can be brought to see this he will condemn his own conduct in forcible language. But if I wish to produce in him equally real conviction of sin in the matter of housebreaking itself, I cannot set about in this way. The only chance of success lies in extending the circle of his interests and sympathies so that he may acquire new dominant aims with which the life of a burglar is incompatible. When I say that he ought not to rob houses I mean that the progressive development of his practical interests, conjoined with the maintenance of systematic unity of aim, would lead him to a stage at which he would condemn robbery as inconsistent with his permanently dominant ends. He means the same thing himself when, without inward conviction based on direct insight, he makes the external admission that he is doing wrong, if, indeed, be does not merely mean that he has done what is usually said to be wrong or what is illegal.

In this way it would seem that every special kind of practical obligation – including, among others, the obligation to pursue truth – may be expressed in the form of

a statement of what would happen if certain conditions were realised. But here, again, we find that the reduction involves a presupposition. It is based on the assumption that men either do in fact universally and necessarily strive after the systematic unity and the fullest extension of their practical interests, or that they ought to do so. The first alternative is obviously untenable. Accepting the second, our problem presents itself anew in a fresh form. How are we to express the duty of self-realisation in the sense defined as a mere proposition concerning what is, or would be, or might be? I shall not myself attempt to deal with this question, because I have not been able to discover a solution which I can regard as satisfactory. I shall, therefore, content myself with an examination of Mr. Muirhead's mode of treating the subject.

(1) Mr. Muirhead says: "While it is true that duty is prescribed by the mind, it is more important for our present purpose to observe that it is prescribed by the facts as well, and that it never could be prescribed by the mind unless it were prescribed by the facts." He then illustrates his meaning by the case of the gambler weighing *pros* and *cons* before staking his last 10,000 francs, and coming to the conclusion that it is his "duty to refrain." Now I see no reason for disputing Mr. Muirhead's contention that in such cases the "moral imperative ought not to be conceived of as something superimposed upon the facts, or setting itself in opposition to them." But it is clear that duty can be elicited from consideration of the facts only under one presupposition. All depends on the primary assumption that the course of action which a man ought to adopt in any given instance is that which he would adopt, if he adequately apprehended with responsive interest and sympathy all that the act and its alternatives involved in their proximate and remote influence on his own mind and circumstances and the minds and circumstances of others. But this primary principle of obligation is just what requires to be interpreted in terms of what is, or under certain conditions, would be.

(2) Mr. Muirhead goes on to say: "the 'ought' is the 'is' in the making, or the is to be; it is the 'is' of the fact conceived of as living and moving as opposed to the 'is' of the fact at 'rest.'" If this were only true it would constitute a perfectly adequate and satisfactory reduction of moral obligation to a matter of fact. But it is very difficult – though it may not, perhaps, be impossible to find an interpretation of "is to be" which will enable us to identify it with "ought to be." Mr. Muirhead certainly cannot mean that, whatever men in point of fact become or develop into, they *ipso facto* ought to become or develop into. What a thing is to be is probably identified by him with what it tends to become, and this, again, is identified with what it is fitted to become by its own

proper and distinctive nature. In this sense we might say metaphorically that an acorn ought to become an oak, because otherwise it fails to develop the capabilities which belong to it *qua* acorn. The real as distinguished from the metaphorical "ought," is on this view to be found in the development of those capabilities which belong to the proper and distinctive nature of human beings as such, i.e., the development of the unity of reason and of feeling as determined by reason amid the growing manifold of presentations and desires.

But it does not in Mr. Muirhead's statement any more than in mine get rid of an ultimate presupposition concerning what ought to be which remains unreduced to an "is" or a "would be." No answer is forthcoming to the question, Why ought I to realise myself in this way? The difficulty becomes accentuated when we consider that human nature as such is capable of vice and moral degradation as well as their opposites, just as living organisms are by their distinctive nature capable of death, decay, and disease, as well as of life and healthy growth. This criticism applies to Mr. Muirhead's further statement that, "Inasmuch as the real nature of anything is that which it has in it to become, rather than that which it already is, the essence of human relations (the facts of the moral world) and the society which is built out of them, is to be looked for not in man's actual achievements in these respects, but in the end or ideal to which he is progressing."

By way of illustrating the independence of "is" or "is to be," he proceeds as follows: "It is, perhaps, not an exaggeration to say the value of our family, university, estate, and cosmopolitan institutions, looked at from within, does not, strictly speaking, rest upon anything that is real at all, but upon the ideal structure of the ought." Now it is obvious that the *significance* of these institutions is constituted by their relation to the ideal structure of the "ought," but to say that their existence depends on the "ought" seems to involve some confusion of thought. As Mr. Muirhead himself tells us, their existence depends on the loyalty of men to the demand of the moral law, not on the moral law itself. But this loyalty is merely a psychological, historical, or sociological fact.

In conclusion, I may say that I am in general sympathy with Mr. Muirhead's mode of dealing with the question. I think it likely that the problem may be solved on the lines followed by him. For this reason it seems to me the more necessary to sift carefully his proposed solution.

Mr. Alexander's identification of what ought to be with what conduces to social vitality or social health

fails, in my opinion, to account for the ultimate nature of "obligation." I can conceive that it might, under certain imaginable conditions, be the duty of a man to do his best to bring about the dissolution of society. If I saw that the preservation of the race entailed unavoidable and unending preponderance of misery, degradation, and slavery over happiness, culture, and freedom of individuals, would it not be a duty, if possible, to put an end to the race instead of preserving it?

IV. S. ALEXANDER

I INTEND TO SUPPORT that reduction of "ought" to a species of "is," "on the lines of scientific thought," which Mr. Muirhead has declared to be naive, and Professor Sidgwick has declared to "eviscerate ethical thought of its essential import and interest." I am in hopes that these different charges may compensate each other: that the naiveté of my doctrine may excuse its criminality.

The practical importance of obligation has invested that idea with a vivid and almost indefeasible belief in its originality and uniqueness. There is, therefore, all the more necessity to state at once how far any attempt to explain obligation can go. Why it is that, in a case where it is said to be my duty to do anything, I should experience the characteristic feeling of obligation, I can no more explain than I can explain why I see a red object red, or feel angry when I am provoked. I can but verify the fact. All that can be expected of any account of obligation is that we should be able to verify that, when the elements contained in the analysis are present, our experience assures us that the feeling of obligation is present; and, secondly, that no essential element has been omitted. I believe that the critics of the so-called scientific or biological position expect tacitly something more. It may be added that no such explanation as reduces obligation ultimately to the level of a fact can alter either its practical importance or the prima facie theoretical difference which compels us to treat ethics as a distinct branch of knowledge.

I will restate very briefly the account of morality and its obligation which has been impugned. We start from the fact that people have certain desires, which are not merely impulses issuing in action but are directed upon objects consciously entertained. For this reason the object of a desire may be called an ideal object or an ideal; at any rate, all ideals are the more or less complicated objects of more or less complicated desires. Owing to the dependence of individuals in a country upon one another, but especially owing to the natural inclinations, whether springing from affection or fear, which are felt towards other persons, any object of action is a social object. A man need never take into account the abstract

conception of the good of society as a whole, but he takes into account, in his actions, their relation to other persons; or more properly any of his objects involves a relation to other persons. Now everyone who acts for conscious objects acts for ideals; but, as a matter of fact, only certain of such ideals are considered good. These constitute the system of good conduct, and the persons who practice them are called good. Other ideals are called by these persons bad. These good persons, so far as they are good, approve good actions and disapprove bad ones. Good conduct, conduct which it is a duty to perform, corresponds therefore to the existence of certain sentiments on the part of good men, which dispose them towards such objects. Where then is obligation? This again is a feeling entertained by good persons, in so far as they are good. It arises from the fact that any particular action is called for by the whole society of which the agent is a member, or by the whole mass of his other sentiments. It includes the approbation passed upon a proposed course of good conduct, or of disapprobation passed upon a proposed course of bad conduct, but it contains something more, namely, the compulsion which the mass of "active" sentiments (whether as felt by the individual himself or by other members of society) exerts to enforce the doing of the good act. In the case of gambling the obligation to retrain arises from the power which the whole character possesses to repress the proposal to gamble. Obligation is therefore felt only so far as a man is good; the bad man is open only to the compulsion of fear, and his conviction of duty is limited to the theoretical knowledge that the people called good require the performance of certain acts. The compulsion contained in obligation (though it may have been generated in a man by external constraint) arises from his other moral sentiments. Obligation is, therefore, approval backed by the force of the whole character.

The criterion of morality is described as social health, or equilibrium, or adaptation to the surroundings. The evidence on behalf of this criterion need not be presented. But, so far as valid, the criterion states the fact that all action called good is such as implies an equilibrium of persons in society or of desires in an individual's mind. It asserts that morality is the compromise which arises out of the attempt to give free play to every person in society or every function in the individual. It does not assert that good action is predetermined by the idea of equilibrium, but that when people act their conduct tends to an equilibrium, and that such conduct is called moral. It is therefore a comprehensive description of the sum total of good conduct by its most characteristic feature. This criterion differs from that of maximum happiness in that it claims to be a primary description of moral conduct. Given the equilibrium, maximum happiness is included.

So much for the criterion of moral action. The total object of moral action is the conduct described by the criterion. On the other hand, the reason why moral conduct becomes the regular rule is found in a process identical in principle with nataral selection, in virtue of which the sentiments of "good" persons impose themselves upon the whole society. They do so because the ideal which corresponds to these sentiments affords a *modus vivendi* for all. This gives them their strength and secures their preponderance. Their general acceptance proves that satisfaction is found in them. Good persons succeed, not because they are good, but because they desire certain objects which commend themselves to the majority. That they do desire these objects, whereas others desire different objects, is a purely natural fact – a difference of tastes. That they succeed is another natural fact. Their victory settles what that ideal of goodness is which subsequent reflection discovers to be an equilibrium of interests. It is a victory enforced by censure and punishment of recalcitrants, by remorse when they themselves fail. It arises in natural superiority and is maintained by pains and penalties. The imposition of good conduct is the most flagrant instance extant of the tyranny of the majority.

Let me now turn to the objections raised. Professor Sidgwick objects that this account of moral judgment, as the expression of sentiments directed towards particular actions, does not explain why some such judgments are true and others false; or why, when two such judgments conflict, one must be erroneous. I will take the latter part first, because the former part is answered *prima facie* by referring either to the existence of the criterion or to the process by which the criterion is established. As to the second, Professor Sidgwick urges that "one psychological fact cannot conflict with another fact; A's judgment that all gambling is wrong does not conflict with B's judgment that some gambling is right." This is perfectly true so long as A's and B's judgments are regarded as parts of the knowledge of the observing moralist. There is no inconsistency in the existence of these two judgments, because they can be explained by reference to the characters of the different persons. But A's and B's judgments, as they exist in A's and B's minds, are not equivalent to the knowledge in the moralist's mind that they do so judge. A body is pulled in contrary directions by two forces, one of ten the other of twenty pounds. There is no conflict between these facts; but you might as well say that there is no conflict between these forces as maintain that there is no conflict between A's and B's judgments. Whether all mental events are actions I will not ask; but, at any rate, these judgments are judgments about action and issue in action. A's and B's sentiments are forces which do conflict, and in two ways: (1) B's

judgment, if acted upon, leads to consequences which may directly or indirectly be harmful to A; (2) the idea of B's gambling shocks A's sentiments, which operate against gambling, and causes him displeasure. What we mean, in fact, when we say that B's judgment is erroneous, is that it is disapproved by that exclusive society for mutual admiration called good men, against whom the miserable wicked have no defence. Mr. Sidgwick himself points out how, when we inquire into the ground we have for knowing the existence of error, we are driven upon the inconsistency of thought with thought, because the inconsistency of thought and things can only be judged by reference to the thoughts in which we know the things. Why should he then object to the view that badness consists not in the inconsistency of an act with some standard of "duty" which is different from the mere sentiments of a certain class of persons, but in its inconsistency with the mass of sentiments directed towards objects which have arrogated to themselves the title of right?

Mr. Sidgwick goes on to say that, while a reference to the end (as social health or equilibrium) may account for why one judgment is right and another wrong, this test can be applied only on the assumption of the ethical judgment that health or equilibrium is itself valuable. This objection is, of course, equally valid against any other account of the end, like maximum happiness, which does not appeal to some ultimate intuition for its obligatory character. I find it difficult to follow the criticism, because health and equilibrium both describe nothing but a certain distribution of activities in the body, whether the body material or the body moral; and this distribution if attainted is nothing but an event. But the criticism appears to admit of either of two interpretations, both of which are fallacious:

1) It may mean, granting equilibrium to be the end of good action, why ought this end to be pursued? Now the end so described, or however described, is nothing but the sum of such actions as ought to be pursued. To the end, as a whole, "ought" is inapplicable. Obligation applies to any particular action in relation to the end. To ask for the obligation to the end is to ask for the explanation of a tautology. The end represents the sentiments of good men; the obligation to it is the compulsion exercised by these sentiments as a whole upon any particular sentiment which aims at any one part of the end. In other words, obligation is an internal relation between the parts of the end, not a relation external to the end. The only intelligible answer to the question, why I ought to promote equilibrium, is to be found in the process by which the equilibrium is established, for that equilibrium constitutes at once the distinction of right

from wrong and the obligation to do particular acts. We can question the value of the equilibrium as a whole only when we are dissatisfied with it; this must be on the ground that it does not satisfy our desires, or, in other words, falls short of some new equilibrium which we desire to establish.

(2) If the criticism does not mean this, it means that we cannot act for the sake of social health or equilibrium without first asking, why it is right to seek the equilibrium. But the idea of equilibrium is a theory; it is not the object of practice. In practice we desire particular objects, and if we think of the whole of the moral system as an end, we think of the mass of actions comprehended under it. We may upon occasion think of the equilibrium as our ideal, in the same way as we may think of maximum happiness; and doubtless we do so in many cases. But we do so because we have assured ourselves that this is the true criterion of any action, and therefore can be made legitimately our direct object. The objection, in fact, seems to me to confuse the characteristic of the moral end, as viewed by the moralist, with the motives of the agents who pursue it. Whether we act up to a social end already determined, or being dissatisfied with this, endeavour to introduce a new mode of life, in neither case do we, as a general rule, act for the sake of the equilibrium which is characteristic of the standard. We act because we desire certain objects. If these objects are imposed by the existing standard they are approved. If they are acts which modify the existing standard they can only be approved by reference to the new standard of which the reformer has a forecast. The only use of entertaining the idea of social health or maximum happiness is to direct our action wisely to those objects which have the characteristics which experience and reflection have convinced us are proper to any end ever proposed as moral, in the same way as we consciously apply approved science in order to abridge the record of guesses and failures which would be the result if we followed our immediate fancies.

Mr. Stout's objection to the biological view of morality, that it does not account for the ultimate nature of obligation, is put in a different way. We ought, he urges, under certain imaginable conditions, to hold it our duty to bring about the dissolution of society, if the preservation of the race meant more misery than happiness, or meant degradation, such as I suppose might happen if, owing to great climatic changes, we had to adopt a simpler mode of life. I confess I find great difficulty in meeting this consideration. I have said that we should be content with that explanation of obligation which we can verify by reference to our experience, provided no essential element is omitted. And am I to conclude that

an essential element has been omitted on the strength of the obligation which would problematically be felt under circumstances as problematical? The objection is valid not only against the criterion of utility, but against any other criterion whatever which can be adopted in the face of data which are certain. It is no more absurd to be bound to secure a moral vitality which consists in extinction, than a greatest happiness which cannot be felt. Even if obligation is something *sui generis*, it is, as Mr. Stout would admit, not independent of circumstances, and affects individuals who will enjoy the performance of the conduct. No other idea of obligation would be consistent with the everyday facts of morality. I do not feel, therefore, that the puzzle falls specially to me to consider. But I will make some remarks upon it. In the first place, not only is it conceivable, but it really happens, that duty should require the destruction of any particular society or portion of society, or of the form which society takes at any particular time. But this would be our duty only in view of the health of some other larger or different society to which the continuance of the society in question, at any rate in its present form, is hostile. "That man's the best conservative who lops the mouldered branch away." Excluding this case, which accords within the doctrine I support, let me remark how extremely problematical it is whether in the cases Mr. Stout imagines any obligation would be felt at all to destroy the society. Perhaps some individuals might feel such an object their duty. But mistaken ideas of duty are possible under any circumstances. I imagine, however, that persons might help on the dissolution of society not from any sense of obligation, but from fear of the future. Does a suicide take his life from a sense of duty or because he has no courage to continue living? And the whole of existence might destroy itself from sheer weariness without feeling any obligation to do so. Not even if universal suicide (and nothing short of this would be to the point) took place by deliberate agreement, would there be the elements of obligation present as I have defined obligation, even though the act would be approved by the whole race? For the approbation which I have described is meaningless if the action approved is to have no effect on the character of anyone at all. Finally, even if obligation were present, the annihilation which follows the act is as much an attainment of equilibrium as it is of happiness or consciousness of duty. But if we turn from these somewhat vain discussions to the guidance of actual experience we find that any excess of pain over pleasure is not an incentive to destruction, but to removing the source of pain, and that when for any reason men's condition becomes simpler or degraded they adapt themselves to the new state, like fishes, which, having to live in the dark, live on, but lose

their sight. Moreover, the theory itself throws light upon the circumstances imagined. For it shows that people do not first consider whether actions, which will secure moral equilibrium, will elevate or degrade character, but, acting for the objects they desire, they secure with it equilibrium and in general elevation, though possibly, in some cases, degeneration. And as to the case imagined when misery exceeds happiness, though it is easy to assert this, it is impossible ever to prove it, and the theory I am supporting, by using the fact that life-preserving acts are pleasurable, maintains that morality produces an excess of happiness if only for the *a priori* reason that morality is the rule of conduct which survives. It secures excess of happiness because those whose feelings would turn the scale in favour of misery, that is to say, who cannot find their place in the social equilibrium, are extirpated.

Mr. Stout explains in a manner with which I am in full agreement, how both truth and goodness imply the self-consistency of their parts. This is nothing more nor less than is implied with respect to conduct in the idea of equilibrium. But he goes on unaccountably to declare that in both cases the reduction of truth and goodness to a fact depends on two assumptions: (1) that all men necessarily pursue truth and goodness, (2) that truth and goodness are worth pursuing. Neither of these assumptions seems to me to be made in fact. The determination of what is called a body of truth as opposed to falsehood, merely implies that truth is that body of knowledge which succeeds in commending itself, and stigmatising other knowledge as false. The theory assumes only that people strive to understand or even merely to apprehend the world with which they are brought into contact, without asserting that they pursue truth as such in any sense. If so much be not granted, if men do not according to their lights seek to understand, the constitution of truth and the recognition of its claim would be unintelligible. As any theory of goodness takes for granted the fact that persons desire objects, so any theory of truth takes for granted that persons apprehend objects. The second supposed postulate has no existence whatever. Truth arises out of the conflict of the various apprehensions of objects, as goodness out of the conflict of desires. The value of truth and goodness is the result of the process which distinguishes true from false and good from bad. The men who seek knowledge, or who seek goodness, need never think of the value of that consistency of apprehension or of desires which constitutes truth and goodness. Or if they do so it is only because they import with their operations an idea which they have derived from reflection on the general character of the process on which they are engaged.

On Mr. Muirhead's paper I have only a few short remarks to make. So far as he maintains that "ought" is in close connection with the Facts of life, arises out of them and again passes into them, I am, of course, in agreement with him. The desires whose equilibrium constitutes morality are desires which are determined by the conditions in which persons have to act. So far as Mr. Muirhead goes beyond this, I am unable to think that his demonstration of how the ought reacts upon the "is," helps us in the solution of the problem before us, and when he goes on to declare the "is" to be dependent on the "ought" because this "ought" represents the real nature of human action, and perhaps of science, he is, I think, com-mitting the error of confusing the order of fact, of nature which always proceeds by efficient causa-tion, with the order of significance. The ideal may be more significant than the fact, but in so far as the ideal works, it works only as a fact – as an idea in the mind of the person who possesses the ideal.

IS THE DISTINCTION BETWEEN "IS" AND "OUGHT" ULTIMATE AND IRREDUCIBLE?

Commentary by Frank Jackson

FRANK JACKSON is Emeritus
Professor at The Australian
National University. His
publications include:
*The Philosophy of Mind
and Cognition* (with
David Braddon-Mitchell,
Blackwell, 1996), *From
Metaphysics to Ethics*
(Oxford University Press,
1998), and *Language,
Names and Information*
(Wiley-Blackwell, 2010).

Commentary by Frank Jackson

I.

THREE THINGS STRUCK ME as I read these papers by Henry Sidgwick, J. H. Muirhead, G. F. Stout, and S. Alexander. I expect many will have the same or similar thoughts.

The first is the way themes pursued by the authors relate to themes much debated today. Is it true that no matter how much information we have about a proposed course of action given in non-moral terms, the inference to a conclusion about whether or not the action in question ought to be performed is, in some substantial sense, an extra step? (Sidgwick and Stout, most especially). Are the rightness and wrongness of actions determined in some relatively straightforward sense by the relative values of available alternatives? (Sidgwick). How is an intuition that some action ought to be done ('feeling of obligation') related to whether or not it should in fact be done? (Alexander, most especially). To what extent should we seek to make sense of morality within a framework that views human being as natural parts of the natural world? (Alexander, most especially, but also Sidgwick and Stout). How does normativity in general, including the normativity of belief, connect to normativity in the sense of morally right conduct? (Muirhead, most especially).

The second is the way the issues are pursued without explicit (*explicit*) notice of the questions they raise in metaphysics and epistemology, philosophy of action, philosophy of mind and philosophy of language. The third is that the supervenience of the moral on the non-moral is never mentioned, not even by implication.

In my commentary, I will try and spell out how some of what our authors say intersects with the matters they neglect, or better, with the matters we would today insist need to be addressed explicitly. I will start with the questions in metaphysics and epistemology and later move on to those in the philosophy of mind and of language. For reasons of space, I will not draw any connections with issues in the philosophy of action.

I will presume a fair degree of familiarity with our texts. That way I obviate the need for an excessive

amount of direct quotation. And, of course, I am going to be highly selective.

II.

Sidgwick writes in ways that invite a cognitivist construal. This is true of all our authors but is especially prominent in Sidgwick. (Of course, they are writing before the various forms of expressivism and prescriptivism in ethics were on the radar.) Sidgwick talks of 'the objects of thought to which the predicates "is" and "ought" are respectively appropriate...' (p. 28) (i. e. he presumes the predicates in question are alike in having *extensions*); he uses the terms 'quality' and 'character' in connection with what is good (p. 28 and p. 29, resp.); he talks of the world of duty as 'a subject of human *knowledge, no less* than the world of fact' (p. 29, my emphasis); he talks of the 'fundamental assumption ... that some such judgments [i. e. moral judgements] are true and others false' (p. 29); and so on. Indeed, it seems to me that it would be hard to make sense of his contribution outside a construal of this kind. But this means that Sidgwick is committed to ethical properties. In particular, he is committed to the view that some acts have the property of being morally right, and others of being morally wrong.

Here the talk of properties is not to be understood in a sense that commits to controversial theses in analytic ontology familiar to us all under such headings as Platonism and Nominalism. Sidgwick is committed to moral properties in the everyday sense in which we are all committed to, e. g., the property of being a table: the sense we capture by talking of ways things might be, and noting that tables differ from, e. g., chairs and from nations in a sense that outruns the fact that they are different objects; or, to say it drawing on the philosophy of language, he is committed to predicates like 'is morally right' and 'is my duty' applying or not applying to actions in virtue of the nature of those actions, much as 'is a table' applies just to those objects which are of the appropriate nature (the nature that, e. g., chairs don't have).

III.

What are the implications of making explicit the metaphysics behind Sidgwick's paper? An immediate consequence is that we need to scrutinise the way he expresses the issue that is on the table. He says, 'The question then is raised, whether this distinction between what is and what ought to be is ultimate and irreducible?' (p. 29) But of course, given his (implied) metaphysics, what

ought to be is part of what is. When we say that murder is wrong, we are making a claim about the nature of murder in the same way we do when we say that murder is on the increase. Again, Sidgwick contrasts the world of duty and the world of fact, but the world of duty is part of the world of fact, given his metaphysics. Arguably, what he should have said is something like the following: 'We can distinguish the way things are that we capture without using ethical terms from the way things are that we capture using ethical terms. Our question is, 'How is the second way things are related to the first?' Indeed, this way of thinking of the issue on the table is suggested by the way that the topic of the symposium is stated in terms that mention rather than use the words 'is' and 'ought'.

On this way of thinking of the issue on the table, it is an issue in metaphysics, albeit one we use a distinction between kinds of terms to identify. For the issue concerns the properties, not the words that we use for them. And for any cognitivist in ethics this has to be the focus, or so it seems to me. When I am told that some action is wrong, aren't I being given putative information about the action, and the putative information of interest isn't that the word 'wrong' in English applies to the action? What is of interest is the property that the action must have in order for the word to apply to it.

What then can we say about the properties in question, and, more particularly, what can we say about the properties that connects with the papers in the symposium?

The first thing to say is that Sidgwick will be right when he says that 'moral judgments ... are a department of psychical fact, and we may analyse and classify them as such and investigate their causes, just as we should do in the case of any other psychical fact...' (p. 29). For instance, to take an example that figures in the symposium, the judgement that all gambling is wrong will be the belief that all gambling has the property of being wrong, and we can discuss the belief's causes and effects much as we can discuss the causes and effects of the belief that the sun is a long way away. Moreover, Sidgwick will be right when he says (p. 29) that the judgement that all gambling is wrong conflicts with the judgement that some gambling is right. He doesn't say it this way, but I take his point to be that their contents are inconsistent, and he will be right because the two judgements assign inconsistent properties to gambling. However, it is far from clear that Sidgwick's more general remarks in this context, directed against the kind of position represented by Alexander in this symposium, are correct. The reason goes back to an issue prominent in recent,

or reasonably recent, philosophy of mind and language.

IV.

Let me spell this out. I will make a start by going back to Sidgwick's remarks at the beginning of the paragraph from which I have just quoted. Here he expresses agnosticism about whether or not the 'distinction between what is and what ought to be is ultimate and irreducible' (p. 29), but goes on to affirm that no reduction 'can be attained by considering moral judgments from a psychological or sociological point of view, as elements in the conscious life of individuals or communities or races' (p. 29). The reason he gives for this claim in what he says following this passage isn't entirely transparent, or at least it wasn't to me (or to Alexander, if it comes to that), but I take the general thrust of it to be that collecting information about the circumstances in which people make moral judgements, although a perfectly proper thing to do (something we have already noted Sidgwick says), cannot explain the way the moral judgements can 'conflict'. This is because people can come to different judgements about, e. g., gambling, and 'One fact cannot be inconsistent with another fact' (p. 29). The idea seems to be that aggregating information about facts that can co-exist cannot explain inconsistency.

The problem for this line of argument is that it neglects the issue of how the contents of beliefs and words get determined. It is a contingent, a posteriori matter that we mean what we do by our words, including our ethical terms. We might have used 'morally good' for things that are square, and 'square' for things that are morally good. We need, therefore, an answer to the question of what makes it the case that they mean what they in fact mean – an answer that will draw *inter alia* on empirical information about word usage in circumstances. Likewise, we need an answer as to what makes it the case that our beliefs, including of course our moral beliefs or judgements, have the contents they in fact have. Any fan of multiple realisability in the philosophy of mind (and there are many) will insist that the answer must draw on empirical information about the functional role of the beliefs framed in terms of inputs, outputs, and interconnections with other mental states, and, on some views, selectional history. The detail isn't important here. What is important is that the answer will draw, in one way or another, on empirical facts about our moral beliefs: what triggers them, what they trigger by way of behaviour, the inferential relations they stand in, maybe their selectional history, etc. (*Mutatis mutandis* for the meanings of ethical terms and sentences.) That is to say, the answer will draw on the kind of information that Sidgwick declares to be irrelevant and the kind Alexander

appeals to in his contribution.

Moreover, once we have to hand an account of how the contents of the various moral judgements are determined – what makes it the case that they have the contents they do have – we will have an entry point into the question of the nature of those contents and, thereby, the identity of moral properties. For the content of the judgement that gambling is wrong, for the cognitivist, is that gambling has the property of being wrong. The reductionist will hope that the delivered identity is with some property describable in non-moral terms.

I emphasise here that I am not supporting the particular answer that Alexander gives. The point I am making is that certain empirical facts about moral beliefs, including their causal and functional roles, and the corresponding facts about moral words, are relevant to the topic of the symposium. For they are relevant to the contents of our moral beliefs and our moral vocabulary and, thereby, to the properties we ascribe when we say that some act is, say, our moral duty. But that's not to say that the empirical facts that Alexander appeals to are relevant; moreover, there is the complication that Alexander focuses on what he calls the 'feeling of obligation' (p. 40) rather than on belief, an issue we discuss near the end.

v.

I have just said that the reductionist hopes that investigations into the content of moral beliefs will tell us that the properties they ascribe are identical with properties ascribable using non-moral terms, but of course it isn't that simple. Once we make the metaphysics behind Sidgwick's approach explicit, it becomes transparent that there are two different ways of being a reductionist. On one, the focus is on properties. To be a reductionist is to hold that moral properties are identical to non-moral properties – not in a sense that would violate Leibnitz's law, but in the sense that the properties picked out by moral terms are a proper sub-set of those picked out in non-moral terms. I will follow a reasonably standard practice and call the latter 'natural properties', but I emphasise that we are giving the notion of a natural property a wide interpretation. We aren't restricting natural properties to physical properties in the sense that figures in discussions of physicalism about the mind. For example, the kinds of properties of phenomenal mental states that dualists insist are missing from the physicalists' world view count as natural properties in the sense relevant to the debate over reductionism in ethics.

On the other way of being a reductionist, the focus

is on one or more of words, propositions and concepts. To be a reductionist, on this conception, is to hold one or another thesis that affirms the existence of a priori entailment relations running from non-moral to moral propositions, non-moral to moral concepts, or non-moral to moral words, or even that there are a priori true biconditionals connecting moral terms with non-moral terms (a view I in fact hold), or some such.

It seems to me that the first way is the way of most interest. My reason is one we noted earlier. We take the trouble to produce words like '(morally) good', 'bad', 'right' etc. not because of the words they are *per se*, but because of what they *signify* – the properties their use attributes to actions, states of affairs, policy decisions, etc. So the question of most interest is the nature of the properties, and it is that question that's addressed by reductionism construed as a doctrine about property identity.

Why don't our symposiasts make explicit the difference between the two ways of understanding reductionism? I suspect (this is a guess) that they hold, maybe without explicitly acknowledging it, that questions of property identity and questions of a priori equivalence of corresponding predicates are, at bottom, one and the same. Many today agree, many disagree. What is common ground today, I take it, is that the issue calls for explicit discussion.

Be all this as it may, let us return to the question of reduction understood as a thesis about the identity of moral properties.

VI.

Why might one share the hope of being able to identify moral properties with natural ones? Two reasons, or two we will mention here; one is familiar in current debates in epistemology, the other concerns the metaphysics of properties.

The first is that it would be nice if we could allow ethical properties to play causal roles. That way we would have some kind of reassurance as to how it might be that our beliefs about the moral properties of, say, gambling, the words that come to our mouths containing words like 'Gambling is wrong', might be reliable guides to the instantiation of these properties. We could think of the beliefs and the words as causally constrained by the properties we hope they give information about. I suspect (but this is another guess) that some of our symposiasts (but not Alexander, I take it) would argue that it was some kind of conceptual confusion to suppose that the instantiation of an ethical property might

make a causal difference. In any case, the reductionists who identify moral properties with natural properties know how to make a start on the epistemology of ethical properties: make sure their candidates to be the various moral properties are properties whose instantiations impinge on us. Matters are famously less clear for anti-reductionists. Again, I am not taking sides (here) on this question. I am simply flagging the topic as one that is prominent in current discussions of the relationship between the moral and the natural, and of reductionism in general, but which is largely buried in the papers under discussion.

The second reason, the one that concerns the metaphysics of properties, connects with the supervenience of the moral on the non-moral. It is plausible that moral properties are *grounded* in the following sense: it is impossible to instantiate a moral property without instantiating a natural property. Supervenience tells us, moreover, that duplication in natural properties ensures duplication in moral properties. It follows that everything with a moral property has some natural property that determines its having the moral property in question. Moreover, both grounding and supervenience are plausibly necessary truths, so plausibly the determination in question is a matter of necessity. (Here, incidentally, is a place where it is particularly important not to read 'natural property' as 'physical property'. It is not a necessary truth that anything with a moral property has a physical property. It is possible that there exist virtuous non-physical angels and evil non-physical demons.) The reductionist's challenge is, How could a natural property necessitate a moral property unless the moral property in question were itself a species of natural property? Again, I am not taking sides (here) on this question, but flagging it as one that calls for explicit discussion. (Although, as we noted earlier, our symposiasts do not mention supervenience, not even by implication, I think it is implicit in their discussion that they accept grounding. Sidgwick, in other places, gives universalisability a prominent role without using the term as such; this suggests that he would support supervenience.)

VII.

A few paragraphs back, we noted that one way of thinking of Alexander's contribution is as offering an account of how ethical terms and beliefs get their contents in terms of functional roles, potentially serving as a precursor to arguing that we can identify moral properties with natural properties. We might also note that his talk of finding an equilibrium gives his presentation of the idea an added contemporary flavour. However, there is an important respect in which his presentation

runs against contemporary wisdom in the philosophy of mind.

Alexander introduces his approach by making much of 'the characteristic feeling of obligation' (p. 40), and talks of it as something we experience, mentioning seeing things as red and feeling angry as suggestive parallels. His idea, in the broad, seems to be that what we need in order to give a complete account of morality is a suitable account of the origins of these feelings. When he says, in a critical tone of voice, that 'critics of the so-called scientific or biological position expect tacitly something more' (p. 40), he is, I think, urging that all that can reasonably be asked for is the right account of where these feelings come from. Balancing the desires of different members of society, being conducive to overall well-being, being selected by pressures coming from the desires of the majority, being an advantageous adaptation etc. – the distinctive parts of his view responsible for its being standardly described as a species of evolutionary ethics, and mentioned by him in summary form on p. 41 – are that which explains the existence and importance of these feelings.

This invites the thought that Alexander isn't really an advocate of evolutionary ethics at all, contrary to received opinion and the opinions of the other contributors to the symposium. He is a kind of subjectivist who adds to his subjectivism an evolutionary account of the causal history of the feeling of obligation and of moral sentiments in general. That indeed was my first thought on reading his contribution. However, there is a way of glossing his view that makes it a version of evolutionary ethics. It is to insist that the causal history he offers should be read as a view about what it takes, or is necessary for, a feeling of approval to be a feeling of *moral* approval. Just as a certain causal history is necessary for a coin to be genuine, or for a child to be the legitimate heir to the throne, so being the product of a certain kind of process of selection (or however precisely one should summarise the many points he makes on and around p. 41 concerning the origins of the feeling of obligation, of moral sentiments etc.) is necessary for a feeling of obligation, say, to be a feeling of moral obligation. Although Alexander doesn't express his view in these terms, I think he might well have found them congenial.

VIII.

The big problem with Alexander's view lies elsewhere, it seems to me. Most (not all) contemporary philosophers of mind will struggle to identify the feeling he talks about. It will be as elusive as the 'feel' of believing that snow is white. A person who believes that snow is

white isn't thereby in a distinctive sensory state in the way that they are when something looks red to them or they are in pain. Likewise, judging that some course of action is one's moral duty isn't to enter into a state with a distinctive feel. One will typically be aware that one has made the judgement, and may be aware that one is or is not disposed to act on one's judgement, but there's no distinctive feeling, or so it seems to me and to many.

IX.

I close with a puzzle for our reading of Sidgwick's underlying metaphysics.

I said near the beginning that he is committed to moral properties. This means he has to embrace a style of correspondence theory for the truth of ethical sentences: a sentence of the form 'X is morally right' will be true just if X has the property of being morally right. The same goes for ethical belief and thought: the belief/thought that X is morally right will be true just if X has the property of being morally right. (In saying this, we are not endorsing a correspondence theory of truth for ethical *propositions*, or indeed, for any propositions; that is a separate and much more controversial matter.) And a position of precisely this kind in the case of thought is suggested by the following words of Sidgwick's, '...from a philosophical or epistemological point of view ... we regard the world of duty and the world of fact alike as objects of *thought*, and – real or supposed – *knowledge*, and discover similar *relations* of thought in both, relations of universal to particular and individual notions and judgments, of inductive to deductive method, etc. ... .' (pp. 30-31, my emphasis)

The puzzle is that, shortly after these words, Sidgwick says, 'True, if we adhere to common sense, the fundamental difference remains that the distinction between "truth" and "error" in our thought about what is, is held to depend essentially on the correspondence, or want of correspondence, between Thought and Fact; whereas, in the case of "what ought to be", truth and error cannot be conceived to depend on any similar relation except on a certain theological view of duty [which he goes on to reject] ... .' (p. 31) Why does he think this is part of common sense? Is he really saying what he thinks himself and calling on common sense as support? In any case, he has, in my opinion, no choice but to embrace a correspondence view for ethical thoughts (and sentences).

PRACTICAL DUALISM

by E. E. Constance Jones

E.E. CONSTANCE JONES
(1848 – 1922) was born
in Herefordshire. She
studied and worked at
Girton College, Cambridge,
eventually becoming the
mistress of the college.
She retired in 1916. Jones
is now best known for
her work in logic, and is
thought by some to have
anticipated distinctions
(e.g. between sense and
reference, and between
meaning and denotation)
more famously attributed
to Frege and Russell, the
latter of whom was her
junior contemporary at
Cambridge. Jones's work
in ethics was less original,
but she was a notable
and important expositor
and defender of Henry
Sidgwick, publishing
a number of articles
elucidating elements of
his work (including the
one published in this
collection), and seeing
through the posthumous
publication of the sixth
and (definitive) seventh
editions of his *Methods of
Ethics*. She also wrote on
metaphysics, and was an
influential and inspiring
teacher to many early
twentieth century British
philosophers.

by E.E. Constance Jones

"Reasonable self-love and conscience are the chief or superior principles in the nature of man; because an action may be suitable to this nature, though all other principles be violated; but becomes unsuitable if either of those is."

Butler, *Sermon III*, end.

I HAVE ADOPTED AS MOTTO Bishop Butler's statement of what Professor Sidgwick calls the "Dualism of the Practical Reason," because, although the form in which Butler states it is not precisely that which I wish to support, he was, I believe, the first thinker who definitely formulated such a dualism. We find, of course, in Plato demonstrations (1) that the Good is Virtue and (2) that the Good is Pleasure; yet these demonstrations are not complementary but contradictory, setting forth what were for Plato opposed views of the ultimate Good.

The doctrine which I wish to set forth and to support is briefly indicated in Butler's words. There are, he holds, two supreme principles of human action, *both* of which we are under a "manifest obligation" to obey. These are, in his view, Reasonable Self-Love and Conscience. For "Conscience" I would substitute Professor Sidgwick's emendation: "those among the precepts of our common conscience" which we "really see to be ultimately reasonable." This substitution leads to the acceptance of Rational Benevolence as the second of the two "chief or superior principles."

My reasons for bringing up for discussion at the present time Sidgwick's doctrine of Practical Dualism are (1) that this doctrine seems to me to be of unique value in the theory of conduct, the present ethical and political situation of the civilised world tending to emphasise this value; and (2) that to the best of my knowledge the doctrine is to a large extent ignored, repudiated, or misunderstood by writers on ethics and politics. This may seem strange in view of Butler's reputation as a moralist, and the fact that Sidlgwick's *Methods of Ethics*, in which Practical Dualism is expounded and maintained, has gone through eight editions between its publication in 1874 and the year 1914, in which the last edition appeared. There has been a steady demand for the book, showing that students of moral philosophy cannot do without it. No doubt there is an explanation of this state

of affairs, but to go into this is not my present purpose, which rather is to put before the reader as briefly as possible the full case for Sidgwick's doctrine of the "Dualism of the Practical Reason," and the way in which he arrives at it.

In the view of "Common Sense Morality," good conduct for an individual consists in obeying certain definite hard and fast rules, without regard to consequences. No doubt we all as respectable members of society acknowledge that we ought to do justice, to be courageous and temperate, to speak the truth, to act with benevolence, loyalty, and gratitude, to keep our promises.

In all this consists virtue as commonly understood. But when we come to reflect upon these rules of action, we find that they are sometimes tautologous, sometimes vague, sometimes inconsistent with one another. To every rule exceptions are allowed. "The common moral axioms are adequate for practical guidance, but do not admit of being elevated into scientific axioms."

The search for rules which can be accepted as scientific axioms leads the inquirer from that Common Sense Morality known as Dogmatic Intuitionism to another phase of Intuitionism, which has received the qualification of "Philosophical." This method accepts "the morality of Common Sense as in the main sound," but "attempts to find for it a philosophic basis, which it does not itself offer; to get one or more principles more absolutely and undeniably true and evident, from which the current rules might be deduced, either just as they are commonly received or with slight modifications and rectifications."

Sidgwick finds such principles in Kant's Categorical Imperative: Act so that thou canst will the maxim of thy action to be law universal; and in Clarke's Rule of Equity and Rule of Love or Benevolence. The Categorical Imperative gives the *form* of a law or general rule. The Rule of Equity is that "whatever I judge reasonable or unreasonable that another should do for me, that by the same judgment I declare reasonable or unreasonable that I should do for him."

The Rule of Universal Love or Benevolence is that "every rational creature ought in its sphere and station, according to its respective powers and faculties, to do all the good it can to its fellow creatures, to which end Universal Love and Benevolence is plainly the most certain, direct and effectual means." The maxim of Prudence or Rational Self-Love is that "one ought to aim at one's own good on the whole." These maxims are "practical principles, the truth of which, when they are explicitly stated, is manifest."

"Most of the commonly received maxims of Duty – even of those which at first sight appear absolute and independent – are found when closely examined to contain an implicit subordination to the more general principles of Prudence and Benevolence, and ... no principles except these, and the formal principle of Justice or Equity, can be admitted as at once intuitively clear and certain, while, again, these principles themselves, so far as they are self-evident, may be stated as precepts to seek (1) one's own good on the whole, refusing all seductive impulses prompting to undue preference of particular goods, and (2) others' good no less than one's own, repressing any undue preference for one individual over another."[1]

If we now turn back to the question "What is the ultimate Good for man?" which Plato answered sometimes by saying "Virtue" and sometimes by saying "Pleasure," it would appear that the answer: *The Good is Virtue*, is excluded. "For to say that 'General Good' consists in General Virtue – if we mean by Virtue conformity to such prescriptions and prohibitions as make up the main part of the morality of Common Sense, would obviously 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 General Good. Wisdom is insight into Good and the means to Good; Benevolence is exhibited in the purposive actions called doing Good" – and so on.[2]

What, then, is the ultimate Good, since it is not Virtue, seeing that Virtue involves a reference to the Good otherwise determined? Is it Pleasure; or, if not, what else is it among "the objects that men have held to be truly Good or the Highest Good"? ("Good" = desirable or reasonably desired). Sidgwick discusses this question in Ch. IX, Bk. I, of *The Methods of Ethics* and in Ch. XIV, Bk. III, and reaches the conclusion that "nothing, is ultimately good except some mode of human Existence," and that "in the view of Common Sense, beauty, knowledge and other ideal goods are only reasonably to be sought by men in as far as they conduce either (1) to Happiness or (2) to Perfection or Excellence of Human Existence."

We have seen that to say Virtue is the Good involves a logical circle, and it is not in accordance with common sense to regard minor gifts and graces, or mere subjective rightness of will as constituting ultimate Good. It follows that nothing can be accepted as ultimately good except desirable consciousness, and this again must be either (1) Happiness or (2) objective relations of conscious minds to, e.g., Truth or Beauty or Freedom. Reflection and reference to common sense lead us to prefer

1. *The Methods of Ethics*, pp. 391, 392, 7th edn.

2. *Loc. cit.*

the former of these alternatives, namely, Happiness, as the ultimate Good. Thus the hedonistic or eudaemonistic end is admitted, and Virtue interpreted as conduct conducive to that end, the principles of Rational Benevolence and Rational Self-Love being accepted as supreme and co-ordinate.

We have, accordingly, reached the doctrine of the Dualism of the Practical Reason. I will quote here Sidgwick's summary (in the concluding chapter of *The Methods of Ethics*) of the results of comparing Intuitionism and Utilitarianism. "We have seen," he says, "that the essence of Justice or Equity (in so far as it is clear and certain) is that different individuals are not to be treated differently except on grounds of universal application; and that such grounds, again, are supplied by the principle of Universal Benevolence, that sets before each man the happiness of all others as an object of pursuit no less worthy than his own; while, again, other time-honoured virtues seem to be fitly explained as special manifestations of impartial benevolence under various normal circumstances of human life, or else as habits and dispositions indispensable to the maintenance of prudent or beneficent behaviour under the seductive force of various non-rational impulses. And although there are other rules which our common moral sense when first interrogated seems to enunciate as absolutely binding; it has appeared that careful and systematic reflection on this very Common Sense, as expressed in the habitual moral judgments of ordinary men, results in exhibiting the real subordination of these rules to the fundamental principles above given. Then, further, this method of systematising particular virtues and duties receives very strong support from a comparative study of the history of morality, as the variations in the moral codes of different societies at different stages correspond, in a great measure, to differences in the actual or believed tendencies of certain kinds of conduct to promote the general happiness of different portions of the human race; while, again, the most probable conjectures as to the pre-historic condition and original derivation of the moral faculty seem to be entirely in harmony with this view. No doubt, even if this synthesis of methods be completely accepted, there will remain some discrepancy in details between our particular moral sentiments and unreasoned judgments on the one hand, and the apparent results of special utilitarian calculations on the other; and we may often have some practical difficulty in balancing the latter against the more general utilitarian reasons for obeying the former; but there seems to be no longer any theoretical perplexity as to the principles for determining social duty."

But we have still to consider the relation of Univer-

salistic to Egoistic Hedonism. "Even if a man admits the self-evidence of the principle of Rational Benevolence, he may still hold that his own happiness is an end which it is irrational for him to sacrifice to any other; and that therefore a harmony between the maxim of Prudence and the maxim of Rational Benevolence must be somehow demonstrated, if Morality is to be made completely rational. This latter view," Sidgwick adds, "appears to me, on the whole, the view of Common Sense; and it is that which I myself hold."

It will be remembered that in Mr. Sidgwick's view no complete reconciliation between Universalistic and Egoistic Hedonism can be demonstrated either on the basis of experience or sympathy, or on theological or metaphysical grounds. He points out, however, that, even so, ethical science is in no worse position, as regards its foundations, than natural science is.

Sidgwick's arguments for the acceptance of the principles of Rational Benevolence and Rational Self-Love separately and on their own merits seen to me to be convincing, and the case for the principle of Rational Benevolence (which aims at the happiness of other human beings generally) to be overwhelming, whether as reached by an examination of Common Sense Morality or as an immediate intuition. It does not seem open to denial that (as Clarke declares) "every rational creature ought in its sphere and station to do all the Good it can to its fellow-creatures" (for Clarke, Good = Happiness). That the Rule of Benevolence is fundamentally equivalent to the "Golden Rule" and the "New Commandment" of the Gospels is also a point in its favour.

I think, however, further, that it is possible to deduce from it alone the maxim of Prudence – that the agent's own happiness on the whole is a reasonable end of his action. According to the principle of Benevolence, we ought to promote the happiness of others – to accept their happiness as our end. But we can only accept it on the ground that each individual's happiness is *to him* ultimately and intrinsically valuable.

Now, a man cannot experience, cannot directly know, any happiness but his own. It must, therefore, be on the ground that *his own* happiness is *to himself* ultimately and intrinsically valuable, valuable in itself, that he can logically regard the happiness of others as ultimately and intrinsically valuable to them. His reasoned belief in the value for others of their own happiness must be based, it can only be based, on his recognition of the value for himself of his own happiness. "It is only if my own consciousness tells me that my happiness is *for me as an individual* intrinsically worth having, only on this condition is there valid ground for holding that the hap-

piness of others is that which for *their* sakes it is worth while for me to promote. Why should I think that another's happiness is any good to him, unless I feel that my happiness is good to me? Can I judge his consciousness except by my own?"[3] All distress at the pain of others, all hatred of cruelty, all indignation at the "injustice "of undeserved suffering, the irrepressible demand that the "wages of virtue" should not be "dust," that there should be a heaven for the good, is based in the last resort on our apprehension of the intrinsic value of Happiness, and this, as we see, must start from the individual's apprehension of the intrinsic value of his own happiness to him. But Benevolence, love of others, is as natural as love of self (as Butler has maintained); and, chronologically, the impulse of Benevolence is often prior to reasonable Self-Love, and a man's own greatest happiness may often depend on the happiness of others, and his acutest misery be caused by the suffering of others. Yet, still his happiness and misery, whatever the *cause*, are *his* in a sense in which they cannot be anyone else's, nor anyone else's his; and the Happiness of any Community can be nothing but the Happinesses of its individual conscious members, untransferable in every case. Mr. Sidgwick has somewhere remarked that "twenty dull people do not make a brilliant dinner party," and applying this idea to the case in hand, we may say also that twenty million unhappy souls cannot constitute a happy nation.

It would thus appear that Benevolence implies Self-Love, and Rational Benevolence irresistibly leads us back to the rationality of Self-Love as our starting-point. Similarly, in the precepts "Do unto others *as ye would* (reasonably would) *they should do unto you*," "Love your neighbour *as yourself*," it is implied that the love of self is logically prior to, and sets the standard for, love of our neighbour.

If we start with that "chief or superior principle" which is Rational Benevolence, we *have* also that other "chief principle" which is Rational Self-Love. If there is a contradiction between these two principles, it is a contradiction which is implicit in the single principle of Benevolence itself. Thus it would seem that the "Utilitarian" or Universalistic Hedonist is, as such, a Practical Dualist; and not only so, but also the man who accepts, broadly, the morality of common sense, is a Practical Dualist in embryo; if he develops logically, he must become a Practical Dualist.

Practical Dualism is, I think, the only ethical doctrine which perceives and fairly faces the claims, for the individual agent, of both self *and* others. It gives a clue to the mixture of good and evil in men – it does not leave us hopelessly puzzled either by domestic shortcoming or

3.   *Proc. Arist. Soc.*, 1903-1904, p. 37.

foreign atrocity. Self-Love and Benevolence are recognised as both natural, and both rational. What seems to be a not uncommon procedure in this connexion is that people professedly and theoretically recognise only, or at least emphasize only, the claims of *others*, as "Duty," "Conscience," "Virtue," the "Ten Commandments," the "New Commandment," and so on; but, in practice, more than redress the balance by a disproportionate attention to self-interest – which people are always ready to attribute as a motive. This tends to confusion of ethical theory, complete psychological muddle, and practical inconsistency.

For a practical dualist the principle of Rational Benevolence is of fundamental importance in politics, that is, "the consideration of the ultimate end or Good of the State, and the general standard or criterion for determining the goodness or badness of political institutions."[4] The relation between ethics and politics is very close – in Sidgwick's view they are in fact parts of one whole – the science of conduct. "On the one hand," he says, "individual men are almost universally members of some political or governed community; what we call their virtues are chiefly exhibited in their dealings with their fellows, and their most prominent pleasures and pains are derived in whole or in part from their relations to other human beings: thus most of those who consider either Virtue or Pleasure to be the sole or chief constituent of an individual's highest good would agree that this good is not to be sought in a life of monastic isolation, and without regard to the well-being of his community; they would admit that private ethics has a political department. On the other hand, it would be generally agreed that a statesman's main ultimate aim should be to promote the well-being of his fellow-citizens, present and to come, considered as individuals; so that the investigation of the particulars of this well-being must be an integral part of Politics."

Politics, indeed, is all-embracing: it aims at "the happiness or well-being of humanity at large." All other ends, such as Freedom, Wealth, Science, are subordinate to this. Politics is, for the most part, consciously "utilitarian," and if common sense morality were not unconsciously aiming also at the General Happiness, our ethical and political action would be even more inconsistent than it is, our ordinary ethical and political thought even more confused.

In ethics the individual agent has to combine or co-ordinate the point of view of Self-Love and the point of view of Benevolence. The statesman *as such* is not embarrassed by the dualism in the same way as the private individual may be. "He exists primarily for the good

4. Sidgwick, *History of Ethics*, 1910, p. 3.

of the governed in the political community to which he belongs. The promotion of the good of his community (with, of course, a due regard to the good of the larger whole of which it is a part) is his *raison d'être*" (compare Sidgwick's *Essays on Public Morality* and *Morality of Strife*).

"But for him, too, the dualism is, from a different point of view, momentous. Though as statesman he is not liable to be faced with the conflict (which emerges primarily as a conflict of motives) between Interest and Duty, between the Happiness of self and the Happiness of others, yet since the community which he administers consists of individuals who are one and all liable to this conflict, it is his business to reconcile the conflict to the utmost of his power, to make it for the interest of individuals to do that which, if they would do it, would be for the Good of the Whole – to furnish at any moment motives sufficiently strong to induce individuals at that moment to do what is for the General Good. In proportion as the attainment of Happiness for self and the attainment of it for others are – so far as the power of government extends – made coincident in any community, in proportion as they are promoted by the same course of action, in that proportion is the community well organised and well governed, to that extent do the members of the community enjoy what Kant calls the 'Supreme Good'; they are both virtuous and happy. The great problem for rulers, as for teachers, is to promote this coincidence of Well-doing and Well-being. Herein lies much hope for the future – the reduction for the individual agent of the conflict between Self-Love and Rational Benevolence does seem to be, to a very considerable extent, in the power of rulers and educators."[5]

5. Article "Henry Sidgwick," in Hastings' *Encyclopedia of Religion and Ethics*.

And a similar hope is, perhaps, possible in regard to the region of international politics. "It would be a great gain," Sidgwick says, "if the whole of civilised society could be brought under a common government, for the purpose of preventing wars among civilised men." He thought (in 1891) that it would be hopeless to aim at this, but urges recourse to arbitration, and, in the case of armed conflict, the impartial imposition on both parties of "rules limiting the mischief of war."

All these, and other, devices for improving international conditions are founded on that maxim of Rational Benevolence which furnishes one of the two "chief or superior principles in the nature of man," and among the most influential means at the disposal of the statesman for carrying its recommendations into effect, here as elsewhere, is the judicious use of the other "chief principle," Rational Self-Love.

What I have attempted to do in this short paper is:

1). To exhibit Sidgwick's view of the Dualism of the Practical Reason and the way in which he reaches it.

2). To state the view which I have formed of a relation between the "two chief or superior principles in the nature of man," namely, Rational Benevolence and Rational Self-Love, according to which it appears that Rational Benevolence implies or includes the Rationality of Self-Love.

ON JONES'S 'PRACTICAL DUALISM'

Commentary by Brad Hooker

BRAD HOOKER is Professor
of Philosophy at the
University of Reading. His
Ideal Code, Real World
(OUP, 2000) developed
a rule-consequentialist
theory of morality. Since
then, Hooker has mainly
been replying to criticisms
of that theory and trying
to develop a new theory
of fairness. However, he
has also published papers
on contractualism, moral
demandingness, normative
reasons, the elements
of well-being, moral
particularism, virtue ethics,
anti-theory in ethics, the
Divine Command Theory,
and the buck-passing
theory of goodness.

ON JONES'S
'PRACTICAL DUALISM'

Commentary by Brad Hooker

1. See the preceding reprint of Jones's article.

E. E. CONSTANCE JONES'S ARTICLE 'PRACTICAL DUALISM' contains far more quotation than most good journals nowadays would tolerate.[1] Jones quotes mostly from her former teacher Henry Sidgwick. But there are also places where she quotes from previous publications of hers on Sidgwick's moral philosophy. Jones's article is so much quotation that pinpointing the article's original contribution is difficult. I will eschew trying to pinpoint originality and focus on the article's main ideas, whether or not they are original.

Although the article touches on the question of whether utilitarianism can be reconciled with common-sense moral intuitions, the article is mainly about the idea that there are 'two supreme principles of human action, both of which we are under a "manifest obligation" to obey.' (p. 64) One is the principle of Rational Benevolence and the other is the principle of Rational Self-Love. Sidgwick concluded his *Method of Ethics* by acknowledging that, unless there is a God to ensure that maximizing impartial good is compatible with maximizing one's own good, the practical realm is founded upon a dualism of principles that contradict one another. The main contention in Jones's article is that 'Rational Benevolence implies or includes the Rationality of Self-Love' (p. 72). I shall explain how there is one reading of this contention that makes it undeniable but other readings that make it implausible. The end of my discussion will comment on Jones's remarks about the relation of utilitarianism to common-sense moral intuitions.

Jones quotes Sidgwick as holding that Rational Benevolence prescribes that, in Clarke's words, 'every rational creature ought in its sphere and station, according to its respective powers and faculties, to do all the good it can to its fellow creatures.' (p. 65, quoting from Sidgwick 1907: p. 384) Since she repeats this formulation at on p. 68, we have to assume she really means it.

However, consider a case where one could benefit one's fellow creatures a little but only at immense cost to oneself. An example might be a case where only by starving oneself to death could one provide some small comfort to others. Obviously, a requirement to make immense sacrifices of one's own good for the sake of small gains to one's fellow creatures would be ridiculously ex-

cessive. In the face of this sort of counterexample, no one would defend a requirement to do all the good one can for one's fellow creatures no matter how small the benefit to them is in comparison to the cost to oneself. Thus, we should be reluctant to construe Rational Benevolence as committed to such an utterly self-denying requirement.

Sometimes, however, Jones seems to take Rational Benevolence to be merely a requirement to intend to promote the good of others.

She writes, 'the principle of Rational Benevolence … aims at the happiness of *other* human beings generally' (p. 68, italics added). And, 'According to the principle of Benevolence, we ought to promote the happiness of others—to accept their happiness as our end.' (p. 68) Jones clearly has in mind a requirement to aim at the good of others equally, counting the happiness of any other human being as no more or less important than the same level of happiness of any other. This conception of Rational Benevolence says nothing whatsoever about how to balance the good of others against one's own. *A fortiori*, it doesn't go so far as to require that one do all one can for one's fellow creatures even when the costs imposed on oneself far outstrip the benefits to others.

What Sidgwick really meant by Rational Benevolence was equal concern for the good of all, including oneself. Jones quotes Sidgwick's remark that Benevolence tells each 'to seek others' good no less than one's own, repressing any undue preference for one individual over another.' (p. 66, quoting from Sidgwick 1907: p. 392) And, 'the principle of Universal Benevolence … sets before each man the happiness of all others as an object of pursuit no less worthy than his own.' (p. 67, quoting from Sidgwick 1907: p. 492)

Equal concern for the good of all of course requires concern for the good of oneself, since one is within the scope of 'all'. Jones identifies 'the maxim of Prudence or Rational Self-Love' as 'one ought to aim at one's own good on the whole' (p. 319). Now it is undeniable that Rational Benevolence *understood as concern for the good of all* includes Self-Love *understood as concern for one's own good*. So there is a reading of Jones's contention that 'Rational Benevolence implies or includes the Rationality of Self-Love' that makes this contention true but unsurprising.

However, if we follow Jones in taking Rational Benevolence to be concern for *others* (in contrast with concern for *all including oneself*), then there is a debate to be had about whether Rational Benevolence is incom-

patible with the rejection of Self-Love. Is there something irrational about caring about the good of others but not about one's own good?

Jones writes,

> I think … it is possible to deduce from it [Rational Benevolence] alone the maxim of Prudence—that the agent's own happiness on the whole is a reasonable end of his action. According to the principle of Benevolence, we ought to promote the happiness of others—to accept their happiness as our end. But we can only accept it on the ground that each individual's happiness is *to him* ultimately and intrinsically valuable. (p. 68)

So she thinks Rational Benevolence is grounded in the 'intrinsic value' of the happiness of others. She then supposes that, if others' happiness has intrinsic value, then one's own happiness must too. So just as the intrinsic value of others' happiness grounds Rational Benevolence, the intrinsic value of one's own happiness grounds Rational Self-Love.

Sometimes her reasoning seems be that Rational Benevolence makes sense only if Rational Self-Love makes sense. She writes, 'Why should I think that another's happiness is any good to him, unless I feel that my happiness is good to me?' (p. 69) and 'Rational Benevolence irresistibly leads us back to the rationality of Self-Love as our starting point.' (p. 69) Here I think she meant 'our apprehension of the intrinsic value of Happiness … must start from the individual's apprehension of the intrinsic value of his own Happiness'. (p. 69) In other words, since one's own happiness has intrinsic value and thus is a worthwhile object of concern, then others' happiness must also have intrinsic value and thus must also be a worthwhile object of concern.

Now Sidgwick's Rational Egoism holds that, all things considered, one ought always do what is best for oneself, even when making a small sacrifice would somehow result in enormous benefits for others. According to Rational Egoism, 'his own happiness is an end which it is irrational for him to sacrifice to any other.' (Sidgwick 1907: p. 497) The 'ought' in 'one ought to do what is best for oneself' is meant to be a *conclusive* (i.e., *decisive, overriding, all-things-considered*) 'ought', not a *pro tanto* one.

Just as Rational Egoism's 'ought' was meant to be a decisive one, so was the 'ought' of Rational Benevolence. Sidgwick's 'dualism of practical reason' was in fact a battle between two principles that contradict one another. Rational Egoism insists that, all things considered, one ought always do what is best for oneself even when one could do greater good for others. Rational Be-

nevolence insists that, all things considered, one ought always do what is best for all considered equally, even when this wouldn't be best for oneself. These claims cannot both be true. Had Rational Egoism and Rational Benevolence both been formulated in terms of *pro tanto* 'oughts', we would have had potential conflict; we wouldn't have had contradiction.

Jones does not note the distinction between *pro tanto* 'oughts' and conclusive 'oughts'. That distinction, W. D. Ross's great contribution to normative ethics, was made roughly a decade later (Ross 1930: chapter 2).[2] So let us consider whether Jones's arguments work given the distinction between *pro tanto* 'oughts' and conclusive 'oughts'.

2. For discussion, see Hurka 2014a: section 3.2.

Her argument was that, just as the intrinsic value of others' happiness grounds Rational Benevolence, the intrinsic value of one's own happiness grounds Rational Self-Love. Does the intrinsic value of others' happiness ground Rational Benevolence? Remember that Jones meant by 'Rational Benevolence' equal concern for the good of *others*. So the question for us to consider is: does the intrinsic value of others' happiness ground equal concern for the happiness of others?

The intrinsic value of others' happiness could ground equal concern for the happiness of others only if the intrinsic value of others' happiness were equal. Maybe it is equal. But Jones does nothing to show this. Very many people believe that it isn't equal. They believe that some people deserve more happiness than others and that some don't deserve any happiness at all. Deserved happiness, the thought goes, has more intrinsic value than undeserved happiness (and deserved unhappiness less disvalue than undeserved unhappiness).[3]

3. See, e.g., Feldman 1997, Kagan 1999 and 2014, and Hurka 2011: ch. 9.

Even if for the sake of argument we assume that the intrinsic value of every other person's happiness is equal, we might still refuse to accept that we must have equal concern for the good of others. Suppose that a stranger and my mother are equally happy. Then the intrinsic, i.e., non-relational, value of this stranger's happiness is the same as the intrinsic value of my mother's happiness. But whether my concern for the stranger should be the same as my concern for my mother depends not just on the intrinsic values of their conditions but also on whether or not I have a special relation to, or connection with, them. Even where the intrinsic value of other people's conditions are equal, I might not be required, and maybe not even allowed, to have equal concern for those people, since some of them might be my friends or family members and others might have no special connection with me.

4. I think Jones was indeed making those assumptions, which are in keeping with utilitarianism's ascription of fundamental importance to happiness but not to desert or to personal connections.

Admittedly, if we assume away the objections raised in the previous two paragraphs, i.e., if we assume that the intrinsic value of others' happiness is equal and that others have no, or equal, connection with me, then we can conclude that I ought to have equal concern for the happiness of others.[4] So much for the grounding of Rational Benevolence, understood as a *pro tanto* 'ought' to care equally about the happiness of others.

But we do not yet have grounding for Rational Benevolence understood as a decisive 'ought' to care equally about the happiness of others. Suppose that the intrinsic value of others' happiness is equal and that others have either no or equal connection with the agent. Well, there might be something else with intrinsic value, such as fairness or justice. And this *might* have more value than others' happiness. If that were the case, then, though the agent ought to care equally about the happiness of others, this 'ought' is not necessarily conclusive. Therefore, in order to have a grounding for Rational Benevolence understood as an decisive 'ought' to care equally about the happiness of others, we have to assume not only (a) that the intrinsic value of others' happiness is equal and (b) that others have no, or have equal, connection with the agent but also (c) that there is nothing else with as much intrinsic value as others' happiness.

We have been discussing the proposition that the intrinsic value of others' happiness makes it the case that one ought to care equally about the happiness of others. First, I pointed to considerations that get in the way of accepting that proposition. Then I pointed out that, even if the proposition is accepted, there is a further question whether the 'ought' is a decisive one. Now I want to add that, if this 'ought' to care equally about the happiness of others is a decisive 'ought', then Jones is struck trying to argue that, because one ought to care overridingly about the happiness of others, one ought to care about promoting one's own happiness.

Of course there is the old point that a degree of self-preservation and even self-development will typically be necessary if one is to be in a position to do good for others. But the Rational Self-Love that Jones wants to defend is not of this instrumental kind. Jones wants to show that Rational Benevolence includes concern for one's own good as a *non-instrumental* end in itself. But it just is not true that an *overriding* concern for the well-being of others 'implies or includes' a non-instrumental concern for one's own good. Moreover, there remains the contradiction between Rational Benevolence and Rational Egoism that so troubled Sidgwick: an *overriding* concern for the impartial good conflicts with an *overriding* concern for one's own good.

As promised, I will close with some observations about Jones's remarks about the relation of utilitarianism to common-sense moral intuitions. Common-sense morality consists of rules requiring honesty, justice, loyalty, benevolence, etc. But 'when we come to reflect upon these rules of action, we find that they are sometimes tautologous, sometimes vague, sometimes inconsistent with one another. To every rule exceptions are allowed.' (p. 65) Because of such objections, Jones follows Sidgwick in thinking that we should look for practical principles that are clearer and more certain and are without exception. The problem with this line of argument, however, is that some of the objections on which it is based have little weight and the others are undermined by the distinction between *pro tanto* oughts and decisive oughts.

The objections that have little weight are the ones about tautologies and vagueness. Most common-sense moral requirements are not tautologies, and common-sense morality can easily survive their being cast aside. Vagueness certainly is a problem, but vagueness can hardly be eliminated entirely.

Admittedly, the rules of common-sense morality can come into conflict with one another. When they do come into conflict, e.g., in a case where honesty and kindness pull in opposite directions, are both rules absolutely binding, providing decisive 'oughts'? Some people think the answer is at least sometimes yes. These people believe in tragic dilemmas, cases where each of the choices an agent could make is, all things considered, morally wrong. But, in at least many cases of conflict, either one rule seems more important in the circumstances than the other, or one of the rules seems to need an exception so as to avoid such conflicts.

5. See Hurka 2014b.

If the rules of common-sense morality are taken to identify *pro tanto* requirements rather than decisive, i.e., all-things-considered, requirements, then the pressure to think they all must be mistaken dissipates.[5] *Pro tanto* requirements are not impugned by occasions when they should be overridden by other considerations. And thus if we think that even requirements of honesty, justice, and loyalty can be rightly overridden when complying with them would result in disasters, we should not assume that this compels us to accept Rational Benevolence, Universal Hedonism, or as we now call it act-utilitarianism. In some cases, protection of aggregate well-being trumps other moral considerations. Those who admit this need not accept that protection of aggregate well-being trumps other moral considerations in all cases.

The utilitarian argument of Sidgwick's that seems to

me best is the argument that we find implicit in common-sense morality the acceptance that we should evaluate codes of rules by their utility. That idea can be developed in a rule-utilitarian way, but such ideas go well beyond Jones's article.

REFERENCES

Feldman, F. (1997). *Utilitarianism, Hedonism and Desert*. Cambridge: Cambridge University Press.

Hurka, T. (2014a), *British Ethical Theorists from Sidgwick to Ewing*. Oxford: Oxford University Press.

_____. (2011), *Drawing Morals*. Oxford: Oxford University Press.

_____. (2014b), 'Sidgwick on Consequentialism and Deontology: A Critique', *Utilitas* 26 pp. 129–152.

Kagan, S. (1999), 'Equality and Desert' in L. Pojman & O. McLeod eds. *What Do We Deserve?* Oxford: Oxford University Press, 1999: pp. 298-314.

_____. (2014), *The Geometry of Desert*. Oxford: Oxford University Press.

Ross, W.D. (1930), *The Right and the Good*. Oxford: Oxford University Press.

Sidgwick, H. (1907), *Methods of Ethics*, 7th edition. London: Macmillan.

THE PLACE OF DEFINITION IN ETHICS

by G. C. Field

GUY CROMWELL FIELD
(1887 – 1956) was
educated at Balliol College,
Oxford, and then taught at
Oxford, Birmingham and
Manchester. He served in
the intelligence department
during the Great War,
during which time he was
for a while a prisoner of
war in Germany. After the
war he taught at Liverpool,
before taking up a chair
at the University of Bristol
in 1926, which he held till
his retirement in 1952. He
published widely in moral
and political philosophy.

by G. C. Field

THIS IS NOT, PERHAPS, ONE OF THE MAJOR
QUESTIONS which the moral philosopher has to face.
But it is of some interest, and of enough difficulty to lead
to certain clear differences of opinion. Everybody is fa-
miliar with the classic argument of Prof. Moore that the
chief notion in Ethics, that of good, is and must be inde-
finable. A somewhat similar argument has recently been
put forward by Dr. Ross and extended by him to the no-
tion of right. Some members of this Society will remem-
ber, also, the lively interchange of opinion which took
place at the last Joint Session, one member demanding
a definition of the term under discussion, while another
denounced this demand as a mere trick of dialectic.

I propose, therefore, to ask the general question sug-
gested by my title. And, in trying to answer it, I do not
propose to begin by giving a definition of "definition,"
and then seeing whether there is any room for the pro-
cess thus defined in ethical thinking. I prefer to begin
at the other end, and ask what processes go on in eth-
ical thinking which might possibly be called definition
or which resemble other processes which everyone is
agreed in calling definition. If we are clear about what
processes can properly go on in ethical thinking, it be-
comes to a large extent a matter of choice whether we
call any of them definition or not. And it will be of help
in this investigation if we look for a moment at the use
of definition in one or two other branches of knowledge.

1. Perhaps the most obvious and typical instances of
definition are to be found in the definitions of Euclidean
geometry. At any rate, it seems probable that this was
the type that Aristotle had in mind when he drew up his
rules for correct definition.

The place of definition in the Euclidean geometry
seems fairly obvious. It is the necessary starting point of
the investigation. We begin with a definition of a figure,
and from it, with the aid of certain axioms and postu-
lates of general application, we deduce other properties
of that figure and its relations to other figures. We thus
have to have a definition to start with. And to arrive at
such a definition is not a very difficult process. The defi-
nition of a triangle, for instance, simply states the most
obvious feature of a kind of figure with which we are all
familiar and of which we have a perfectly distinct idea.

In other cases, though the definition is not very difficult to arrive at, it is not quite so obvious as this. The Euclidean definition of a circle, for instance, is not immediately obvious to the beginner in geometry, though the thing defined is perfectly clear and distinct to his mind. He has to stop and think for a moment before he sees that it is true of that sort of figure.

The difficulty is not very great. But it introduces us to a point that will assume more importance later – that is, the point that we may "have an idea of" a circle, and know what we mean by the term before we know the definition, before, that is, we are aware of the particular feature of the circle that a geometrician takes as the definition of it. This would be more obvious if we considered some of the other definitions that Greek mathematicians attempted of the circle, for instance, that it was the largest area that could be contained by a line of given length. I have heard unmathematical people, like myself, express doubts whether that was true of a circle or not. But, in spite of this, they would not have admitted that they did not know, in some sense, what a circle was.

2. We may consider next the place of definition in zoology and botany. I am referring here to the earlier work of classification of species and genera, which was carried out by the older natural historians. The modem biologist, in general, I suppose, is not very much interested in this. But it forms a necessary basis for his subsequent investigations.

The process of definition here is similar to definition in geometry, in that it involves the statement of the general features which distinguish one species or genus from the others. But the place it occupies in the investigation is entirely different. It is certainly not the necessary starting point of the investigation. It is rather the conclusion of it. And the natural historian does not use his definition as a basis from which to deduce the other properties of the species. Nothing follows from the definition, in the sense in which the conclusions of a geometrical proposition follow.

If it were always necessary to have a clear definition of what we were talking about before starting our investigations on it, it would be difficult to see how the natural historian could ever start his investigations at all. Of course, he has some idea of the things he is talking about, he means something by plant, animal, dog, horse, fish, etc., before he begins to investigate and classify them. Human beings at the pre-scientific stage made some distinctions based on some observable differences between one kind of living being and another. And the first scientists start from that. But there is nothing that we could call a definition. There is only what, for con-

venience, we may speak of as the vague popular idea of dog, horse, etc.

What is contained in this vague popular idea it would be very difficult to say. Clearly it is based on certain obvious visible characteristics. But which characteristics the unscientific man takes in determining whether he is going to call any particular animal a dog or not it would be almost impossible to determine with any certainty. We might arrive at a conclusion by a long and careful psychological investigation. But the important point to note is that such an investigation would be of no interest at all to the scientist. It would tell him nothing of what he wanted to discover. He simply accepts the fact that we do have vague popular ideas of the different kinds of living creature, and all this does for him is to point to the direction in which he can begin his investigations. But, once begun, he carries on his investigations by observation and experiment in entire disregard of the original idea. He looks for and finds facts which are not dreamed of at the pre-scientific stage, for instance, the facts of internal structure, which are generally taken as the most important features in the definition. These features are not in any sense contained in the original idea, nor could they be said to be implied by or deduced from it. When we arrive at a zoological definition we could not in any sense say that this is what we really meant by the term all along. Sometimes, indeed, our definition may contradict the original idea. Most people who know no zoology would probably call a whale a fish, like Herman Melville's whalers, or a spider an insect.

There is one more point that we may raise in passing before going on to our main subject. That is the question what light is thrown by these instances on the statement sometimes put forward that there is one kind of definition which consists in an arbitrary statement of what we are going to mean by a certain term. There is clearly no place for this kind of definition in either natural history or geometry. On the other hand, we need not accept the assumption apparently made by Aristotle that there is one and only one right definition for every general term, and that any other definition is wrong. We find, for instance, that the definition of a circle changes in passing from Euclidean geometry to conic sections. But that does not mean that the more elementary definition is wrong. What it does mean is that there is a certain freedom of selection from among the general properties of the thing defined according to the context in which we are going to use the definition.

This freedom of selection, however, is severely restricted. Most obviously it is restricted by the facts. We can only select among the properties that really belong

there. It is restricted, again, by the context. If we want to investigate in one particular direction, we may choose the definition that will be most helpful for that investigation. But which definition will actually be most helpful is a matter of fact which we have to discover. And we are restricted by the ordinary use of language. Even if we modify the ordinary idea of what a word means in our final definition (e.g., when we define "fish" so as to exclude whales), we must still keep as near to it as possible. We must not define "fish," for instance, so as to exclude all or most of the creatures usually called fishes. To say, " By 'fish' I am going to mean a two-legged animal with feathers," would be entirely pointless. An arbitrary statement of what I am going to mean by a word, if it is really entirely arbitrary, is not a definition, or even a form of definition. It is merely a way of being silly.

It is possible that definition has sometimes been supposed to be arbitrary because it has been confused with a different process, arbitrary naming. We get a certain amount of that in scientific investigations when we invent a new technical vocabulary. Here, however, the usual process is reversed. Instead of beginning with a word and then arriving at a definition, we get our definition of the general kind of thing that we have discovered in our investigations, and then look about for a name for it. Even here, however, there is generally some reason for the choice of names. But the possibility of this process illustrates one important point, namely, that definition is never merely of names, but always of something that the name means to us. Otherwise we could not, as we clearly sometimes do, arrive at the definition of a class of objects before we find a name for it. Of course, more often we begin with a familiar name which already conveys some meaning to us, and so we are apt to speak loosely at times of defining a name or word. And there is no harm in that as long as we realise that it is only a loose and popular mode of speech.

3. We now turn to our main subject, definition in ethics. No one doubts, of course, that some of the general terms used in ethics are capable of being defined, even if some of them are believed to be indefinable.

When we consider the work of the moral philosopher in the light of these analogies, two or three points seem to come out clearly at the outset. One of these is that it is impossible in ethics to start, as geometry does, with any definitions which will be generally and immediately accepted and recognizably applicable to the objects of our study. If we could find such clear definitions to start with there would be no call for specifically philosophical thinking about the subject at all. It would be quite a different kind of thinking that would be required. What

gives rise to the need for philosophical thinking is that we are faced with ideas or notions which, though in common use, are not at all clearly conceived, and therefore not immediately definable. And it is the first, if not the only work of philosophical thinking to make these ideas clear and definite. In this respect the position of definition in ethics is analogous to its position in natural history. It is not the starting point, but the goal of our inquiries. It comes at the end, not at the beginning of our investigations.

On the other hand, there is one respect in which the situation in ethics is quite different from that in natural history. To the naturalist, as we have seen, the vague popular idea of any kind of animal, with which we start before we know any zoology, is of no interest at all except as merely pointing the way at the beginning. He then goes on to study the observable facts, and is not concerned at all to analyse or clarify the original idea. "That's not my idea of an elephant," an unscientific person is reported to have said on hearing the scientific account of this species. To which the zoologist, quite naturally, replied, "Perhaps not, but it is God's idea." What is contained in "my idea" of an elephant in this sense is completely irrelevant to the studies of the zoologist.

To the moral philosopher, on the other hand, what is contained in "my idea" of good or right or justice or selfishness is of vital importance. It forms the main part, if not the whole, of the subject of his investigations. At any rate, it is an essential part and a part which calls for hard and prolonged effort. It seems to me that one of the most frequent causes of error in ethics is that the investigation into the content of the ideas of good, right, etc., which actually are or have been held has not been sufficiently widespread and exhaustive. We have no observable facts, different in kind from these ideas and discoverable by quite different methods, to which we can turn, as the naturalist does, for the real subject of our investigation. The starting point for ethics is always the moral judgments of mankind and what is implied in them, and we can never entirely get away from these as our main source of knowledge.

The technique of this process of clarifying the vague ideas with which we start would be an interesting subject of study. Though many people have attained a considerable degree of success in the process, there is little explicit discussion of the material that we have for it and the methods by which it should be treated. There is room for a new logic, or perhaps a psychologic of ethics. There are also, no doubt, certain psychological and metaphysical difficulties in the suggestion that it is possible to discover more in an idea than those who entertained

the idea were aware of (or had any idea of). The position is not like that of the naturalist, who claims to discover a lot more, and sometimes something quite different, in the fact from what there was in the original idea of it. For here we do not yet know whether there are any facts that we have access to, except by an examination of ideas. But we cannot enter into these problems here. It is an undeniable fact that something of the kind goes on. But there is room for considerable difference of opinion about the correct description of it. I suppose in practice we generally say that we have discovered something more in a person's idea than he knew himself if we find him using it or applying it in a way that would only be justified if this something more was included in it.

There is, however, one point in this connection on which I should like to say a word, because it is concerned with the interesting question of what the qualities are which make a good moral philosopher. It sometimes seems to be supposed that the only virtues which a philosopher needs are what I might call the logical virtues, a sense of form and system, a passion for coherence and self-consistency, a love of precise definition, a keen eye for fine distinctions of meaning, and kindred qualities. No one, I hope, would undervalue these qualities. But, for the moral philosopher in particular, I would suggest that it is a profound mistake to treat those qualities as the sole, or even the chief qualifications necessary for his task. At the risk of being misunderstood, I would go so far as to say that it is possible to be too exact and consistent, or rather, perhaps, to insist on exactness and consistency inopportunely at the wrong time or the wrong place.

At any rate, we must remember that the development of these virtues is, if I may be allowed the metaphor, a question of sharpening the instrument with which we think. And an instrument, however sharp, is of no value unless there is something to cut with it. Moral philosophy, as I understand it, consists primarily in reflection on moral experience and criticism of moral assumptions. And we cannot reflect on these things until we have got some sort of acquaintance with them. It seems to me that there is a quality required in ethics analogous to what we might call a good nose for facts in the natural scientist. And this is a different quality from the capacity for constructing a consistent and systematic theory to explain the facts, and just as essential.

This does not mean, as I think it is sometimes taken to mean, that the moral philosopher should necessarily have an extensive and intense moral experience of his own, that he should be living a life of continual struggle against temptations, that he should be labouring under

a sense of sin, or anything of that kind. Doubtless he must have moral experience of his own, and must take this seriously, not just as an object of idle curiosity. But to be living a life of strong emotional stress and strain, to be going through fierce moral conflicts, would probably be, at the time at any rate, a positive disqualification for reflecting on it. Partly, of course, any strong emotion is unfavourable to reflection while it is being experienced. But what is much more serious, such strong emotional experience would tend to concentrate our attention too exclusively on our own experience, which in the nature of things must be very limited, and pay too little attention to the experience of other people, which is just as essential a part of our data.

The important qualification for a moral philosopher is, therefore, not so much moral experience of his own as a certain sensitiveness and receptivity to the moral experience of other people, and to the moral assumptions or ideas that are taken for granted by the people around him. He must be able, in some way, to let these enter his own mind in the form of assumptions or vague ideas, so that he may be able, then, to make them explicit and interpret and criticize them. We may add that he should be able to do this not only for the ideas current around him, but also for the ideas current in other ages. It is very doubtful whether the highest levels of ethical speculation are attainable without a historical sense.

This is the point at which the possible dangers of premature clear thinking arise. Current moral ideas and assumptions are necessarily vague and inexact, and probably often self-contradictory. But even the contradictions and confusions are part of the data for reflection. We must be able to receive them into our minds in that form before we begin to clear them up and make them definite and self-consistent. If we begin our work of cutting them down on the ground of self-contradiction too soon, if we begin limiting their meaning in order to get at a clear definition too hastily, we may easily find that we have rejected without proper examination some of the most essential parts of our data.

I seem to have wandered from the place of definition in ethics to the place of definiteness in ethics. But it is not all irrelevant, for we may presume that definition is one form of definiteness. At any rate, we have got so far that the first task of the moral philosopher is to make clear and explicit the vague general ideas that are held about the objects with which he deals. He has to discover as much as he can of what is implied by the ways in which the chief moral notions are or have been used. This is a sort of definition. But it is not a sort of definition with which he can rest content. For, as has been

suggested, he will find when he has made these implications explicit that they by no means always coincide. Sometimes, indeed, they seem flatly to contradict each other. They, therefore, have to be subjected to progressive examination and criticism, until we finally reach an answer, that satisfies us to the main question of ethics. This main question may be formulated thus: What sort of facts must we suppose there to be in order to account satisfactorily for human beings having these ideas about them?

Thus the whole of this stage of our investigations consists, in a very real sense, in arriving at definitions. This is not, perhaps, the whole of ethics. For when we have got these results we ought to be able to draw certain conclusions from them by a process of deduction. But it is certainly the most important and the most difficult part. So the place of definition in ethics appears to be co-extensive with the greater part of its field. Further, it is a continuous and progressive process. We have no grounds, except as a matter of temporary convenience in a particular part of the investigation, for picking out one feature as being, in a special sense, the definition. Even in geometry we saw how the sharp Aristotelian distinction between the essence, given in definition, and the properties broke down. And in ethics it is even more obviously untenable.

These considerations will indicate the attitude that should be taken up to the suggestion that we must give definitions before our investigations start. As applied literally to ethics the suggestion is obviously absurd. Definitions are the conclusion of the process, and cannot be demanded at the beginning. To the person who says, "You must state precisely what you mean by your terms before you can discuss them," we must reply, "It is only by discussing them that we can find out what we mean by them."

On the other hand, I think it is possible, on occasions, to carry this refusal to give preliminary definitions too far. At least it ought to be possible to give, if not a definition, a preliminary indication, a kind of sign-post pointing in the direction of the thing we are going to investigate. It might be done by citing one or two typical instances. In fact, this is often the most satisfactory method, though not the only one. But we must always remember that it is only a preliminary indication. The reason why some people are frightened of going even as far as this is that they feel that once they have committed themselves to any such statement, they are bound to hold by it for the rest of the discussion, and that if they restrict or enlarge or modify it in any other way they are thereby convicted of inconsistency and self-contra-

diction. But in reality it would be a very unfruitful discussion if it did not produce considerable modifications and developments of the notions with which it started.

We may suspect forgetfulness of this truth when we find textbooks and courses of lectures giving, as they too often do, an inordinate amount of space at the beginning of their treatment to an attempted definition of their subject. I remember a great teacher, who influenced my early thinking more than any other single person, who was fond of exercising his acute powers of criticism on these definitions. He would take the definition of a subject given, either verbally or in writing, by one of its exponents and proceed to examine his subsequent treatment of his subject in detail, and show how far he departed in practice from his original definition. As young men we were enormously impressed by these devastating criticisms, and I think we sometimes wondered how any psychologist or economist could bear to go on with the study of a subject which apparently did not really exist. Such criticism was doubtless a valuable warning against taking any preliminary definition too seriously. But beyond that, it seems to me now a singularly unprofitable exercise. I much prefer the method of the Scottish professor, who, after discussing for some time various suggestions for distinguishing between logic and epistemology, concluded, "The only really satisfactory definition of these two subjects that I can give you is that logic is the subject on which I shall lecture to you on Mondays, Wednesdays and Fridays, and epistemology the subject on which I shall lecture on Tuesdays, Thursdays and Saturdays." We must hope, however, that by the end of his course the relations between the two had become fairly clear, though probably not expressible by any simple formula.

There remains for examination one question, which to some will seem the most interesting, if not the only interesting part of the discussion. Under what conditions and on what kind of grounds can we pronounce any ethical notion to be indefinable? It seems, if our previous discussion has any truth in it, that such a statement itself admits of a variety of meanings. We may mean that our idea of it is indefinable, or that we think of it as something indefinable. What we think it to be is, I suppose, the strict sense of the phrase "what we mean" by any term. But we might also mean that, however we have been accustomed to think of it, it is in fact a simple unanalysable quality which can only be named and not described further. This would mean, according to our account of ethical thinking, that the only way to account for our thinking and speaking thus and thus of it is to suppose it to be such a simple quality. Let us consider these possibilities in order.

We start, it is suggested, with a vague general idea, a consciousness of a sort of a something we do not know exactly what, to which we apply the term in question. Now, in one sense, this is necessarily indefinable. Just because it is a vague general idea, it is different from the clear explicit idea expressed in the definition. The two are not exactly equivalent, as the meaning of two synonyms would be. If arriving at a definition means an advance in knowledge, the words of the definition must express something more than was present in the original idea. That is why a definition is a significant statement at all.

This is, of course, quite different from having a clear and explicit idea of a simple unanalysable quality. We have such ideas, as for instance of any particular colour. But it is obvious that we do not have such ideas of moral facts, at any rate to begin with. If we had there would be no possibility of beginning any discussion about them, indeed no room for philosophical thinking in such questions at all.

There is, however, another possibility that comes somewhere between these two. We may find that in our ordinary use of some moral notion, such as good or right, we use it in a way that implies that it is a simple unanalysable fact about which nothing further can be said. I find it rather difficult to imagine what sort of usage could be said to imply this. But, at any rate, it seems to me clear that we do not in fact use it in this way. It seems certain that our actual use and applications of these notions imply a good many things about them which can be put into words. Consequently they cannot be said to imply that they are indefinable.

So we are left with the possibility about any such notion – let us take "good" as a typical example – that, though we often seem to assume it to be something more, in reality it is only a simple indefinable quality.

Here, again, I find it difficult to see what sort of proof there could be of this. There is the well-known argument in *Principia Ethica* which argues that good must be indefinable, because, whenever a definition is attempted, we can always ask with significance of the complex so defined whether it is in fact good. I have never been able to find any plausibility in this argument. It is not clear what sort of significance such a question is supposed to have. It may mean that we can never be quite sure of the correctness of any definition that we offer, that the possibility of its being wrong is always in our minds, and that therefore we can still raise questions about it. This is no doubt sometimes true. Indeed, if the process of definition is as we have described, it seems the right and proper attitude to take. But it obviously does not exclude the pos-

sibility that the definition may be correct. It is also true, however, that even if we are certain that the definition is correct we can still ask the question significantly, in the sense that we can have something before our minds in asking it. For, whenever the definition is not immediately self-evident, we can still retain a memory in our minds of what we originally meant by the term, namely, our first vague general idea. And, as we have seen, the vague general idea is necessarily different from the more precise idea expressed in the subsequent definition. So we can always put a question about their relation to each other. This, however, applies in many cases in which a definition is admittedly possible. Though I know, for instance, that part of the zoological definition of a fish is that it is cold-blooded, I can still attach a meaning to the question, Is a fish a cold-blooded animal?

There is one further point of importance in this connection. If it were proved in any way that calling a thing good meant that it possessed a simple indefinable quality, it still would not necessarily follow that this was an important or interesting fact. It would only become important if it could be shown that it meant nothing more than that, that there was no further fact that could be asserted about all the things we called good. If there was any further fact or group of facts that could be truly asserted about anything we called good and about nothing else, it is obvious that in any argument or statement we could always substitute this for "good" without saying anything untrue. And if it led us on to further knowledge, it would be a much more interesting and important fact than the mere presence of a simple, indefinable quality.

I think we could find a convincing illustration of this argument by considering once more one of our elementary geometrical ideas. It seems to me clear that what we understand by a circle really has a simple indefinable quality, which we apprehend directly. That is what we think of when we first learn the use of the word, and it remains in our minds even after we have learned the various geometrical definitions. We could always distinguish in our minds between this simple quality, which we should probably speak of in crude language as "what a circle looks like," and any of the facts about that sort of figure given in the definitions. So, when we are faced with Euclid's definition, still more when we are faced with the complicated formula by which a circle is defined in higher branches of mathematics, we could truly say, in a sense," That is not what we *mean* by circle: that is a further fact *about* it." But no one would think such a statement very valuable, nor would it be regarded as invalidating the mathematician's right to his definitions. If we are interested in extending our mathematical knowl-

edge, the simple indefinable quality of a circle becomes uninteresting and unimportant.

The distinction, therefore, which Dr. Ross so frequently insists on (e.g., *The Right and the Good*, Chapter I, *passim*, *Aristotelian Society, Supplementary Volume*, X, p. 61), between the attribute which we mean by the term, and the further attribute or attributes necessarily connected with it, seems to me an unreal one. Partly, it smacks too much of the sharp distinction between essence and properties which we are agreed in abandoning. But also it seems to me to misrepresent the nature of ethical investigation. What we mean by "good " (or "right" or any other moral term), in the first place, is the vague indefinite idea with which we start. But this only sets the problem. What we are trying to find is the nature of the facts that we must suppose to exist in order to account for the way in which we think about these matters. And anything that we can say about them may equally be taken as part of their definition, in the only sense in which definition is possible in ethics at all. It may be that these considerations point in the direction of the doubts lately raised by Mr. Joseph (in *Some Problems in Ethics*) whether goodness should really be thought of as a quality at all. But that, I believe, is to be the subject of future discussions.

OBJECTIVITY IN ETHICS AND MATHEMATICS

Commentary by Justin Clarke-Doane

JUSTIN CLARKE-DOANE
is an Assistant Professor
at Columbia University.
He is also an Honorary
Research Fellow at the
University of Birmingham,
UK, and an Honorary
Research Associate at
Monash University,
Australia. He earned his
PhD from NYU in 2011
under the supervision of
Hartry Field. His work
centers on metaphysical
and epistemological
problems surrounding
apparently a priori
domains, such as morality,
modality, mathematics, and
logic.

OBJECTIVITY IN ETHICS AND MATHEMATICS

Commentary by Justin Clarke-Doane

2.   By "geometry", I mean a branch of pure mathematics. Obviously not all consistent geometries are true of physical spacetime. (I also assume that no one geometry is "metaphysically distinguished" or "carves at the joints" in the sense of Sider (2011). I make a similar assumption in Section III.)

SUPPOSE THAT ETHICAL AND MATHEMATICAL CLAIMS are truth-apt. Field (1931) raises an interesting question. How do axioms, or first principles, in ethics compare to those in mathematics? In this note, I argue that there are similarities between the cases. However, these are premised on an assumption which can be questioned, and which highlights the peculiarity of normative inquiry.[1]

I. OBJECTIVITY IN MATHEMATICS

Which is the true geometry? Field sometimes writes as if this is a serious question (Field 1931: 86). But most philosophers and mathematicians today would disagree. There are various geometries – e.g., Euclidean and hyperbolic – each of which is consistent if the others are. Rather than privileging any one geometry, it is natural to hold that all consistent geometries are true (under a face-value Tarskian truth definition). They are simply true of different structures.[2]

By contrast, it is commonly supposed that a "foundational" theory, such as some formulation of set theory, can be false without being inconsistent. ZF + the Axiom of Choice (AC) and ZF + the negation of AC are not generally thought to both be true – like geometry with the Parallel Postulate and geometry with its negation. But they are no less consistent if ZF is consistent. There is supposed to be an "objective" fact as to whether every set has a choice function.

II. ETHICS AND SET THEORY

It is a familiar point that in both ethics and mathematics we seem to "have no observable facts…to which we can turn, as the [empirical scientist] does, for the real subject of our investigation" (Field 1931: 88). But if set theory is "objective", in the sense in which geometry is not, then the analogy between ethics and set theory, in particular, can be carried further.

First, if set theory is objective, then there is a gap between consistency and truth in set theory, just as there is supposed to be a gap between (logical) consistency and truth in ethics. The overwhelming majority of con-

sistent set theories are false, just as the overwhelming majority of consistent ethical theories are false.

Second, if set theory is objective, then set-theoretic axioms seem to be scarcely more "self-evident" than ethical "axioms".[3] Consider the Axiom of Infinity. This says that there is an infinite (inductive) set. *Given that consistency does not suffice for truth* in set theory any more than it does in ethics, how could this be self-evident? Even if it is "metaphysically" necessary that there is an infinite set, it certainly seems *intelligible* that there is not. As Mayberry writes,

> The set-theoretical axioms that sustain modern mathematics are self-evident in differing degrees. One of them – indeed, the most important of them, namely Cantor's axiom, the so-called axiom of infinity – has scarcely any claim to self-evidence at all (2000: 10).[4]

Finally, given that consistency does not suffice for truth, and that few axioms of interest are self-evident, the proper method of inquiry in set theory seems to resemble the proper method of inquiry in ethics – "reflective equilibrium" (Rawls 1971). We identify plausible propositions, and seek general principles – axioms – which systematize them. The latter may pressure us to reject some of the propositions with which we began as we seek harmony between the two. Of course, this process requires determining what follows from what. It is unsurprising that ethics and set theory might proceed via proof *in some sense*. However, if both areas are objective, then we are not just trying to determine what follows from various axioms. We are also trying to determine what axioms are true – i.e., "the facts that we must suppose…in order to account for the way in which we think about matters" (Field 1931: 95). As Whitehead and Russell write,

> The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it (1997: 59).

III. OBJECTIVITY IN SET THEORY

I have argued that if set theory is objective, then there are similarities between ethical "axioms" and set theoretic axioms beyond the familiar one that both seem to be non-empirically justified. But contrary to the assumption of Section I, set theory, and foundational mathematical theories generally, may not be objective. They may be relevantly like geometry. As Hamkins writes,

3. Of course, an ethical particularist may regard the search for ethical "axioms" as misguided. But for the purposes of this article, I assume, with Field (1931), that it is not.

4. See also Boolos (1998: 130). A related point is that set-theoretic reductions may be vulnerable to Moore's Open Question Argument, at least if its premise is that "we can never be quite sure of the correctness of any definition that we offer" (Field 1931: 93). Consider the various set-theoretic reductions of the natural numbers, such as Zermelo's or von Neumann's. Benacerraf (1965) noted that more than one is formally adequate, and that there is no obvious reason to privilege any one formally adequate reduction over all others. He took himself to have thereby showed that the numbers were irreducible. But whereas Moore (2004) concluded that since moral properties are irreducible, they must be *sui generis*, Benacerraf (1965) concluded that since the numbers are irreducible, they must not exist at all (Clarke-Doane 2008: 246, fn. 5). Of course, an objectivist about set theory can be an anti-objectivist about questions on which alternative reductions of the numbers differ.

> [G]eometers have a deep understanding of the alternative geometries, which are regarded as fully real... The situation with set theory is the same....[S]et theory is saturated with [alternative universes]....[S]et theorists [make] the same step...that geometers...made long ago, namely, to accept the alternative worlds as fully real (Hamkins 2012: 426).[5]

How should we understand this view? It is uncontroversial that every consistent set of axioms – set-theoretic or otherwise – has a model. That is the Completeness Theorem, which is itself a theorem of standard set theory. But in claiming that ZF + AC and ZF + ~AC are both true, the anti-objectivist is presumably advocating more than the Completeness Theorem.[6] The view is not that every consistent formulation of set theory has a model built out of some background set theory, but that it has an *intended* model – i.e., that every consistent such formulation is satisfied under a face-value Tarskian satisfaction relation (Field 1998: 333). The intuition is that, just as no one concept of point or line should be metaphysically privileged, no one concept of set should be. (Of course, some such concepts may be more interesting, fruitful, and intuitive than others.)

IV. ETHICS AND SET THEORY AGAIN

If such a view of set theory is correct, then the analogies of Section II break down. First, if set theory is not objective, then no matter what set-theoretic beliefs we had had, so long as they were consistent, they would have been true. If one could argue that we could not have easily had inconsistent set-theoretic beliefs, and that the set-theoretic truths could not have easily been different, then one could argue that our set-theoretic beliefs are *safe* – i.e., that we could not have easily had false ones (Clarke-Doane forthcoming).[7] However, in light of apparently pervasive (non-logical) ethical disagreement, it is hard to see how to argue that our ethical beliefs are safe.

Second, given knowledge of set-theoretic anti-objectivism, the truth of set-theoretic axioms may be more self-evident than the truth of ethical "axioms", because the consistency of set-theoretic axioms may be more self-evident than the truth (as opposed to consistency) of ethical axioms. If anti-objectivism is true of set theory (but not of ethics), then the fact that it is "impossible in ethics to start, as [set theory] does, with [axioms] which will be generally and immediately accepted" is less *prima facie* puzzling than it might otherwise be (Field 1931: 87).

Finally, assuming that consistency suffices for truth in set theory, the proper method of inquiry in set theory

5. See also Balaguer (1998).

6. Though Burgess seems to interpret Balaguer as merely advocating the Completeness Theorem in his (2001), p. 80.

7. This presupposes the radical formulation of anti-objectivism above. One could instead advocate a less radical formulation of the view according to which, while both of ZF+AC and ZF+~AC are true, only one of, e.g., ZF+Con(ZF) and ZF+~Con(ZF) is (despite both being consistent if ZF is). More conservative formulations of anti-objectivism are also possible (Gaifman (2012): section 2.4), as are more radical formulations (Priest (2013)). For the purposes of arguing that our set-theoretic beliefs are safe, it seems sufficient to argue for a conservative formulation of anti-objectivism (since, presumably, we could not have easily believed the likes of ZF+~Con(ZF)).

does not seem to resemble the proper method of inquiry in ethics. The question of whether AC is true is like that of whether the Parallel Postulate is true. Given a determinate use of "is a member of", the question has an answer – and, for all that has been said, it may depend entirely on the way the mind-and-language-independent sets are. But in learning it we are really just learning whether we are talking about this universe of sets or that, rather than learning what universes of sets there are.[8] If set-theoretic anti-objectivism is true, then we *already know* that ZF+AC is true of some universe of sets (assuming that we already know that ZF+AC is consistent). The interesting question is what follows from it and other consistent sets of axioms. In this sense, the proper method of inquiry in set theory may approximate the "Euclidean ideal."

By contrast, since there is supposed to be a speculative distance between (logical) consistency and truth in ethics, it is a considerable challenge to find ethical "axioms" whose truth is remotely uniquely determined by the data points with which the process of reflective equilibrium begins.

V. TRUTH AND NORMATIVITY

If set theory is not objective, then set theory is *in a sense* trivialized. If logic is objective, then the question of what follows from set-theoretic axioms remains genuine.[9] But no peculiarly set-theoretic questions seem to remain genuine. One can ask which set theory regiments *our* concept of set, or satisfies some theoretical or aesthetic desiderata. But given set-theoretic anti-objectivism, there is no question of which "consistent" such concept is satisfied. All are.

Could ethics be trivialized similarly? Imagine that a philosopher convinces us that, contrary to all appearances, ethics too is like geometry – that every consistent ethical theory is true, albeit true of different entities. In addition to goodness, obligation, and so on, there is shgoodness, shobligation, and so on. Indeed, for every logically consistent ethical theory, there are corresponding properties, and all of them are instantiated "side by side".[10] Knowing that there are logically (even if not Kantian) consistent formulations of both deontological and consequentialist ethical theories, we conclude that each is true (albeit of different entities). Is our deliberation as to whether we ought to lie when utility would be maximized thereby trivialized (and likewise for every question on which logically consistent ethical theories diverge)?

It is hard to see how it could be. A general – even if

8. Strictly speaking, we are learning whether we are talking about this universe of set-like-things or that. Given a determinate use of "is a member of", nothing failing to satisfy the axioms true of the corresponding relation will count as a set.

9. How to spell out the claim that logic is objective is not straightforward. (Obviously, we cannot say that logic is objective if not every consistent set of logical axioms is true, since the claim that a set of sentences is consistent is itself relative to a logic.) For relevant discussion, see Beall and Restall (2005) and Field (2009).

10. Hartry Field uses this locution to describe Balaguer's view of sets in his (1998).

not universal – rule is that if we conclude that we ought to X, then we cannot continue to regard the view that we ought to not-X as on a par. But given that that view *is* on a par with respect to truth, learning that "we ought to X" is true seems insufficient to resolve our deliberation. While knowledge that any consistent set theory is true, and knowledge that ZF+AC and ZF+~AC are both consistent, frees us of the question of whether AC, something similar would not seem to hold in the ethical – and, more generally, normative – case. The fact-value gap appears to be even wider than Hume and Moore suggested. Even knowledge of the *normative* facts may fail to resolve a normative deliberation.

REFERENCES

Balaguer, M. (1998), *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.

Beall, J.C. and Restall, G. (2005), *Logical Pluralism*. Oxford: Oxford University Press.

Benacerraf, P. (1965), 'What Numbers Could Not Be.' *Philosophical Review* 74: 47—73.

Boolos, G. (1998), 'Must We Believe in Set Theory?' in his *Logic, Logic, and Logic*. Cambridge MA: Harvard University Press.

Burgess, J. (2001), 'Review of Platonism and Anti-Platonism in Mathematics.' *Philosophical Review* 110: 79—82.

Clarke-Doane, J. (2008), 'Multiple Reductions Revisited.' *Philosophia Mathematica* 16: 244-255.

_____ (Forthcoming) 'What is the Benacerraf Problem?' in F. Pataut ed. *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*. Berlin: Springer.

Field, G.C. (1931), 'The Place of Definition in Ethics.' *Proceedings of the Aristotelian Society* 32: 79 – 94.

Field, H.. (1998) 'Which Mathematical Undecidables Have Determinate Truth-Values?' in Dales, H. G. and Oliveri, G. eds., *Truth in Mathematics*. Oxford: Oxford University Press. pp. 291–310.

_____ (2009), 'Pluralism in Logic.' *Review of Symbolic Logic* 2: 342–

359.

Gaifman, H.. (2012), 'On Ontology and Realism in Mathematics.' *Review of Symbolic Logic* 5: 480—512.

Hamkins, J.D. (2012), 'The Set-theoretic Multiverse', *Review of Symbolic Logic* 5: 416-449.

Mayberry, J. (2000), *The Foundations of Mathematics in the Theory of Sets*. Cambridge: Cambridge University Press.

Moore, G.E. (2004) *Principle Ethica*. Mineola, NY: Dover.

Priest, G. (2013), 'Mathematical Pluralism.' *Logic Journal of the IGPL*. 21. 4—13.

Rawls, J. (1971), *A Theory of Justice*. Cambridge MA: Harvard University Press.

Sider, T. (2011), *Writing the Book of the World*. Oxford: Oxford University Press.

Whitehead, A.N. and Russell, B. (1997), *Principia Mathematica to *56 (Cambridge Mathematical Library)*. Cambridge: Cambridge University Press.

THE LANGUAGE OF POLITICAL THEORY

by Margaret MacDonald

MARGARET MACDONALD
(1907 – 1956) was
educated at Bedford
College, London, eventually
receiving her PhD under
the supervision of Susan
Stebbing (who also
supported her financially
through her studies;
MacDonald came from a
very poor background).
MacDonald went on to
become a fellow at Girton
College, Cambridge,
where she studied with
Wittgenstein and Moore.
In 1938 she moved to St
Hilda's College, Oxford.
From 1948 she edited the
journal Analysis (which
she had helped found in
1933) until her untimely
death during heart surgery
in 1956. She mostly wrote
on philosophy of language
and (especially later in her
life) aesthetics, but also had
a more general interest in
methodological questions
about philosophical
discussion.

by Margaret MacDonald

I HAVE RASHLY CHOSEN A SUBJECT about which I am more than usually likely to talk nonsense. For I know little about either history or politics. But having recently been forced to read a fair amount of what is called political theory or political philosophy, I have become both puzzled and interested by the curious notation in which much of it seems to be written. One meets here a "contract" which one is carefully warned was never contracted; an "organism" unknown to biology; a superior "person" or higher "self" with whom one can never converse; an "association" or "corporation," whose objects are obscure and which is not listed in any of the recognised Directories. All these descriptions, analogies or pictures have been applied to the State. One or other of them can be found in the works of the most notable political philosophers from Plato and Hobbes to Laski and MacIver. Here, too, will be found elaborate discussions and disputes about whether men are or are not "naturally" social; whether they "really" will what they don't will; whether there is a Law of Nature or a "natural law" not established by any known empirical methods; whether freedom or "objective" freedom is not properly judicious coercion in the interests of order, etc.

There is a genuinely philosophic air about these strange uses of ordinary words. They seem to resemble the replies sometimes given to the haunting doubts which attack us when we reflect on other subjects. On our sensible experience, for example. Is it perhaps only a perpetual illusion? Or on moral actions. Can an action ever be completely disinterested? Or on other people. Do they have feelings as we do, or are they merely perfectly acting automata? How can we ever be sure? It seems, then, likely that the tales about the social contract and the unmeetable person will be related to similar puzzles. I do not, however, intend to expound in detail any of the answers in which the words I have given are key words. I shall avoid exegesis of Hobbes or of Hegel. I want rather to discuss how the uses of these words with the pictures or analogies they embody are related to the puzzles by which they were suggested and to the ordinary uses of language about social relationships and political affairs. What sort of propositions are they and how do they function? For, at first glance, they seem very peculiar. To be told you are party to a contract, of which you were

unaware, and which is nothing like what anyone would ever call a contract, seems to have little to do with giving your vote at a general election, sending your child to a State school, or paying a fine for exceeding the speed limit. Nor is your depression at the Labour Exchange likely to be much relieved by being told that you "really" willed your unemployment (you would never have thought so, unaided) or that the State is a very superior moral person, only even more anonymous and inaccessible than the Permanent Secretary to the Ministry of Labour.

One trouble in politics is to determine how far the questions are empirical and to what extent they are linguistic. Another is to discover what are the ordinary uses of the words involved. For many important words used in political discussion have a degree of vagueness which makes it even easier in political than in other branches of philosophy to disguise a linguistic elucidation or recommendation as an important factual discovery. A further problem is the causation of these puzzles. Is it merely philosophical discomfort about language that induces people to ask certain questions about their social life and accept these answers? If not, is this philosophically important?

I said that I had little knowledge of history. It is sometimes said that no one can understand or criticise political theories without a thorough knowledge of history. Hobbes and Locke cannot be properly understood without knowing the history of the English Civil War, the Revolution of 1688 and their relation to these events. Rousseau cannot be detached from the conditions in France immediately before the French Revolution of 1789. Hegel is inexplicable apart from the luscious yet strenuous atmosphere of the Romantic Movement and the beginnings of German nationalism. All these theories arose in peculiar circumstances of crisis in the particular societies of which their writers formed part and cannot be discussed as though they were of general application like the propositions of mathematics. This, however, is not quite true. It is certainly true that the propositions of politics are not like those of mathematics and it does indicate that practical as well as purely philosophical dissatisfactions have frequently co-operated to move philosophers to write political philosophy. Indeed, they have usually done so with the avowed intention of influencing political affairs. Nevertheless, they never supposed themselves merely to be writing tracts for their times. Locke doubtless wished to justify the Revolution Settlement, but not merely by considering how a reasonable social life was possible in seventeenth century England, but upon what relationship the life of the members of any community, divided into rulers and ruled, must

be based if it is not to appear contemptible to rational human beings. What justifies us in forming political societies, in obeying laws, in being subjected to other persons? This is not a puzzle peculiar to any age. Moreover, so far from having died with the controversies of the seventeenth and eighteenth centuries, the "contract theory" is now being revived.[1] But even this may be philosophically as well as historically important. For present political conditions may be somewhat similar to those in which contract theories formerly flourished. Historical circumstances then may be important in answering the question, "Why have philosophers been induced to ask these questions and accept these answers?" It may be objected that this is to confuse causes and reasons. A philosopher may be moved to doubt the existence of matter because he has swallowed too much opium, but this would be completely irrelevant to any reason with which he supported his view. Why, then, should it be philosophically important that he asks certain philosophical questions because he feels oppressed by the government, or, alternatively, because, like Hobbes, he is worried by the lack of order in the country? It is because the circumstances in which they are asked and answered may work differently for different kinds of philosophical propositions in influencing their effects. And this may be connected with their philosophical "point." With some, e.g., those of "pure epistemology," their importance may be negligible; with others, those of ethics, perhaps, and, even more, those of politics, they may be more important.

What must also be considered, then, are the practical and psychological effects of these problems and their answers. No one will deny that in political affairs philosophical nonsense may have serious effects. Is this philosophically relevant, or not? To deny that it is seems to reduce philosophising to mere scholastic verbalism. Not for any moral reason. Not because philosophers ruin themselves and their subject by sitting in ivory towers and talking about the uses of words instead of considering how the people perish (or flourish) on nonsensical theories and slogans. But simply because, not to try to understand how this language has effects, even though it may give no information, is to miss half its philosophical point and so is bad philosophy. The philosophical "point" of a remark (or the "point" of a remark which is of philosophical interest) is, at least partly, connected with the cause or reason which induces people to go on making it, though it can neither be supported nor refuted by any empirical evidence. It may be false, it may, if taken literally, be meaningless, but they feel that it has some use. This does seem to be relevant to the understanding of some philosophical remarks, if not of all.

1. Cf. Gough, *The Social Contract*, 1936. Introduction and last chapter, H.D. Lewis, "Is there a Social Contract," *Philosophy*, January and April, 1940.

It is true that no solipsist refuses to converse with others, unless he is also suffering from incipient schizophrenia. Nor does the sceptic about the existence of material objects sit down very gingerly on every chair for fear it isn't really there. For these problems have not, usually, been suggested by any practical difficulties about communication or knowledge in ordinary life. Nor will any answer which the philosopher gives to them be likely to alter his subsequent behaviour. The problems of epistemology are mainly academic. Their practical causation and effects, therefore, are unimportant. That is perhaps why it is easier to see from such examples the predominantly linguistic character of philosophical problems, so emphasised recently by Wittgenstein, Wisdom and others. They can be traced, roughly, to a certain discomfort which the philosopher obscurely feels about what seem to be unjustifiable inconsistencies in our uses of certain words, e.g., those of "know" and "feel." And once he can be made to realise this and that no linguistic change which he may wish to suggest will give him that super-empirical information about the world which he supposed possible, he will cease asking unanswerable questions. The whole drama might be played by two solitary sages on a desert island. But I am not convinced that Butler, e.g., was merely puzzled about the use or misuse of the word "interest" or that he supposed that such misuse was the only mistake of his opponents. He was worried because their philosophical remark that "All action is really selfish" was seducing many people to a disregard of their duties. And though Burke, quite rightly, thought most of what Rousseau wrote was, strictly, nonsense, he did not underestimate his influence on the French Revolution. For whereas a person would be thought slightly crazy who took seriously his doubts about the uniformity of nature and refused to eat his dinner for fear the laws of nature might have changed since yesterday and it would now poison him, many people would not think it at all absurd that anyone who said, "An action is right only if by doing it the agent will promote his own advantage" should so interpret this as to neglect most of the actions that would ordinarily be called his duties. They might think he would come to a bad end, but not necessarily in Bethlem. Yet, in a sense, to ask whether all action isn't really selfish is a senseless question. And to assert that all action is "really" selfish is to make not an empirical but a grammatical statement. It expresses either a misuse of or an intention to extend the use of "selfish" to cover actions to which it does not ordinarily apply. But the distinctions formerly marked by the ordinary uses of "selfish/unselfish" must reappear in the new notation if it is to fulfil all the tasks of the old. Nor does this change give us any fresh information about our duties to others. In this it resembles

the epistemological puzzle about "know." When clearly stated, it takes the form of a linguistic recommendation. Yet it has, or may have, or perhaps only seems to have, certain effects which have been considered important. Not that it predicts any such effects. A linguistic recommendation predicts nothing about behaviour. It is empty of factual content. How then does it work so as to seem to need taking seriously in practical life? How does anyone "act" on a purely grammatical statement, except in speech and writing? Yet it does seem sometimes as if they do. The connection of utilitarianism with social reform is another instance. So that completely to understand ethical problems and theories, more than linguistic considerations are required. Or rather, perhaps, different sorts of elements may be involved in linguistic considerations.

If this is true of ethics, it is even more true of politics. Consider the statement, "The authority of the State derives from the contract or agreement by which men consented to give up certain liberties, to form a society and submit to government in order to obtain greater benefits. The interests of a State, therefore, are subordinate to the interests of its members." The attitude thus expressed may have importantly different results from the one expressed by "The authority of the State is absolute for it embodies the 'real' will and permanent interests of its members. It does, moreover, further certain historical and/or divine purposes incapable of fulfilment by any or all of its members. The State, therefore, is a moral person of a higher type than its members who must be subordinated to it." There may be a sense in which neither of the theories epitomized in these statements is directly verifiable by the facts. For the "contract" and the "real will" may correspond to nothing directly discoverable. There may be another sense in which all the facts to which both theories appeal are the same for each. There is then no empirical means of deciding between them. But do not two statements or theories mean the same if all their empirical consequences are identical? Yet the "contractual" and the "organic" views of political relations would never be ordinarily said to mean the same and they have had very different effects. If the difference is not an empirical one of finding facts which will support the one and refute the other; if it is not a difference in their truth or falsity, what sort of difference is it? How can they differ in meaning without differing in verifiable consequences? But how do they differ? They differ, obviously, in picturing political relationships with the help of two very different images. One represents them under the guise of a contract freely entered into between responsible agents who understand the provisions and are prepared to keep them unless infringed by the other party. Joining society in general and keeping its laws is rather

like joining a Trade Union and agreeing not to blackleg or work for less than the minimum rates, so long as the Union, on its side, agrees to maintain and improve the conditions of labour. Or it is like undertaking to provide goods or services in return for certain payment. Most people have had some experience of such agreements. They know what they imply and how they feel about them. If they have accepted the terms, they do not resent being bound by them, so long as they are observed by the other contractees. And, if not, either the law will enforce the terms or they will be released from their share of the obligation. They do not regard themselves, after entering such agreements, as being any "higher" or "lower" than they were before. Such contracts are convenient devices to secure desired social ends. They are useful, and their obligations should be respected, but no sensible person would rhapsodize about them. If then, people picture the relation between themselves, the State and the Government in these terms, certain consequences will tend to follow. They will tend to be affected, emotionally as well as intellectually, in some ways rather than others. They will probably tend to stress the fact or the need for the *consent* of the governed to its governors. For no one can enter a contract without consenting to it. They will emphasise the importance of the *responsibility* of governors to the governed. No contract can be solely one-sided. Because of the reciprocal nature of contract, their attitude to rulers will be critical rather than reverential. Certainly nothing done by rulers to fulfil their part of the bargain will be accepted because proceeding from a higher moral authority than that of mere individuals. The attitude induced by the "contract" picture might be expected to stress personal freedom and the existence on sufferance of all governments; to be, in general, liberal, democratic and unmystical. And it has, historically, tended to produce this result. It encourages the view that social arrangements of all kinds are made by men for their own ends and can be altered and even ended at their will and pleasure. This does not preclude acknowledging that some arrangements, e.g., those comprised by the State, are very important, even that they function as fundamental conditions for most others. Only that they are not sacrosanct. Nor does it follow that changes must or will be undertaken without due regard for the customs and traditions of the past and the welfare of the future as well as of the present. But only that if, in spite of all this, they are *consented* to by a majority of the present members of a society, no higher authority can be found with a *right* to prevent them. The "contract" view can take account of every fact stressed by other views, but its own difference of emphasis alters their point or effects. In political theory, as, indeed, in philosophy generally, it is very often not what is said, but the spirit in

which it is said, which makes the difference.

The other picture tells a very different story. My relation to an organism of which I form part or to my "higher" self is not determined by free choice. "The State," said Burke,[2] "ought not to be considered as nothing better than a partnership agreement in a trade of pepper and coffee, calico or tobacco, or some other such low concern, to be taken up for a little temporary interest, and to be dissolved by the fancy of the parties. It is to be looked on with other reverence … It is a partnership in all science, a partnership in all art, a partnership in every virtue and in all perfection … it is a partnership not only between those who are living, but between those who are living, those who are dead, and those who are to be born … linking the lower with the higher natures, connecting the visible and invisible world, according to a fixed compact sanctioned by the inviolable oath which holds all physical and all moral natures each in their appointed place." Compared with this splendid spectacle we, who compose any society here and now, are very small fry indeed. Should we not accept with becoming gratitude the fortunate chance that permits us to abase ourselves before this embodiment of "all virtue and all perfection?" "Consent," "choice," can mean only acceptance of what seems good to it, and not to our unimportant selves. This is the attitude of submission to and reverence for what is done by Authority, especially if the Authority is or represents what is old and respectable – or, now-a-days, if it commands forces of great physical power – which is induced by this language. Consent, freedom, criticism, which the contract picture emphasised, are not necessarily denied, but they are minimised or re-interpreted. I am no foe to liberty, said Burke, but it must be a liberty connected with order. If a man is misguided enough to resist the General Will, say Rousseau and Bosanquet, he must be forced to be free. The individual is trivial; the social organism or organisation is almost sacred.

These two ways of picturing political relationships may, then, have very different practical and psychological effects which may induce people to want to go on using them, although they learn nothing much from them about political affairs. Some people like to feel part of a vast and important organisation in which their chief function is to admire and obey. The picture of themselves deliberating about contracts, making decisions, criticising representatives, is much too fatiguing. For others, any picture in which they were wholly subordinate would be intolerable. A similar situation sometimes occurs in science. It is true, I believe, that all the planetary motions could, with suitable complications, be described as well by the Ptolemaic as by the Copernican system. In one sense, therefore, they mean the same. But that in another

2. Burke. *Reflections on the Revolution in France*.

they do not is shown by the fierce resistance met by Copernicus and his followers. It had very little connection with the scientific value of their theory. The Ptolemaic theory included the picture of man and his world at the centre of the universe with the heavens revolving round them in cosy circles. The alternative Copernican picture terrified people. They felt lost, insecure, unimportant. Something had gone for ever. Yet nothing had gone, for the facts were precisely the same for each. But the point of the two notations, with their accompanying pictures, used to describe the facts was very different. In their psychological effects they were very different theories.

It may be objected that this use of the words "picture and "image" is itself a misuse, or, at least, an extension of the ordinary use of these words.[3] This is true. I cannot draw the social organism as I can a rat, nor paint a portrait of the "higher" self. And though I can imagine the scene when King John signed Magna Carta, I cannot similarly imagine that when Hobbes' pre-social beings contracted to form society. I can have an image of signing a building contract but not of signing the social contract. But, it may be urged, when it is said that the State resembles an organism or is based on contract, no particular organism or contract is meant. What is thus asserted is a general resemblance between political relationships and those between the parts of any organism or any contractees. But this seems very peculiar. I cannot have an image or paint a picture of a general resemblance. The words "image" and "picture" appear to be used for something which cannot be imaged or pictured at all, in the ordinary senses of these words. But this peculiar usage can be recommended. It emphasises the fact that philosophical remarks resemble poetic imagery rather than scientific analogy.[4] There is a pictorial or analogical element in most theories, scientific as well as philosophical. In poetry there are metaphors and images, but no theories. And philosophical theories have always seemed slightly odd, if not bogus. The use of "analogy" suggests "argument by analogy," i.e., the deduction of new verifiable facts from a suggested resemblance which increases our knowledge about the world. Philosophical theories have no such application. Nevertheless, in common with the scientific analogies they have other, psychological and semi-logical effects. Compare, e.g., the different effects of the planetary theories already mentioned or of the "mechanistic" and "purposive" hypotheses in biology. Philosophical remarks about social contracts and higher selves work chiefly in these ways. Perhaps to look for a contract or a new biological entity after reading Locke and Hegel on the State is only slightly less absurd than to look for flaming tigers after reading Blake or to ask how Wordsworth knew "at a glance" that he saw ten thousand daffodils. The theories of the scientist give new in-

3. I owe this point to Mr. G. Ryle.

4. Cf. Wisdom. "Other Minds." *Mind*, Vols. 49 and 50.

formation about empirical facts; they also induce certain emotional and intellectual attitudes. The language of the poet is predominantly emotive; that of the philosopher less so, but both also have some relation to certain facts, though not that involved in the application of a scientific analogy. They do, however, partly by the use of certain images and metaphors express or call attention in a very vivid way to facts and experiences of whose existence we all know but which, for some reason, it seems important to emphasise.[5] I do not wish to say that philosophy is (inferior) poetry and not (pseudo) science, for it is neither, but philosophy. But it is sometimes useful when considering philosophical theories, and particularly political theories, to realise how unlike scientific theories they are, in some respects, and how much, in others, they resemble the works of the poets. Rousseau is far more like Shelley than he is like Lavoisier. The use of the words "picture" and "image" stresses this resemblance and avoids the scientific associations of the word "analogy." This gives some justification for the extended use of these words.

But, according to their authors, political theories profess to explain certain puzzles about social life which must now be examined.

The surprising fact about political life, according to Hume, is the ease with which the many are governed by the few. Why should people thus submit to the jurisdiction of others of their own kind? Obviously, it is not solely because of the constant exercise or threat of physical force to compel conformity. The ruled, as Hume said, are always more numerous and therefore more powerful than the rulers. Not even Hitler can literally turn the whole of Germany into a concentration camp. For if he did, one result would be, presumably, that production would cease, and power would soon be useless. Nor is force any explanation. For it is conceivable that someone should prefer, and many people have preferred, to die rather than obey rulers of whom they disapproved. Nevertheless, most people most of the time do obey laws and accept the control of governments. It must then be either because they want to, or because they believe they ought to do so. But they do not always want to, and sometimes when they do want to they think they ought not to. The fundamental puzzle of political philosophy, then, is to find a valid reason for political obligation. Why should men be obliged to obey laws and be penalised if they do not? This leads to consideration of the nature of that which appears to command and enforce laws, viz., the State. And there is perhaps an even more "fundamental" puzzle. Why should man live with others of his own kind at all ? The laws which political obligation acknowledges are the rules of societies. But is it

5.   I think this is true of some poetry, at least, but I do not wish to dogmatise about the function of poetry.

necessary to form political societies? Is man "naturally" social or only by convention? If the second, why was this convention adopted, and how? We all know how individuals join a trade union, a church, a club. These are particular societies. But how and why did we all join society in general? What sort of process was this?

These questions have the familiar tone. Philosophers do not ask whether Zulus or Dodos exist, but whether any material objects exist: not whether I can climb Mount Everest, if I choose, but whether I can do any action, however carefully I choose. So the political question is not "Why should I pay income tax?" or " Why should I support the present British Government?" but " Why should I obey any law, support any Government, acknowledge the authority of any State ?" Why, indeed, should I be a member of any civil society?

I cannot consider all the answers which have been suggested to all these questions. Two have already been mentioned. If I have contracted with others to form a society and obey certain rules then my justification for keeping the laws will be that I formerly promised to do so, at least as long as they were generally observed by the other parties to the contract. But this theory leads to no original in the facts. When did I sign this agreement and with whom? One answer is, you did not sign it, but your ancestors did, only so long ago that all trace of it has been lost. The original Magna Carta of society has vanished. Can it really be on account of this undiscoverable transaction that we keep the laws of England in 1941? For suppose, after incredible labour, archaeologists found the lost document, should we feel happier about observing the Education Acts? Ah, that settles it. Now we know why we should send Johnny to school at five instead of into the fields to mind the sheep. Absurd, of course. And what are the provisions of the contract? Am I bound by it to observe laws yet to be made, of which I know nothing, just because they are laws of the contracted society and government? That would be a very peculiar contract to sign. No, the existence of a contract of the kind required cannot be verified. And no contract which could be discovered would answer our question. This is always admitted by political philosophers. "The social contract theory is really an attempt at analysing the logical presuppositions rather than the historical antecedents of the State," says Mr. Gough.[6] The answer it gives to the question of political obligation is that our only justification for obeying laws and governments is that we have consented to do so and that political obligation is not an asymmetrical relationship between rulers and ruled. Hence the use of the contract picture in supporting the claims of individuals and groups against despotic governments.

6.   *The Social Contract*, p. 4.

The word "contract" then is admittedly not used in its ordinary legal sense of the State and the basis of political obligation. It is only "as if" we had signed a contract. The contract theory points not to a contract but to the fact that what we mean by saying that we ought to obey a law is that we have consented to it. But, as Hume said, unless habit, indolence and indifference are to be taken as consent, very few laws would be obeyed at all on this criterion. To how many do we individually consent, and how do we do so? However important, this cannot be the whole story.

What then of the view that the laws of the State ought to be obeyed because they are the edicts of some higher being with which each of us is for the time being identified, or because they represent what we ourselves "really" will in our best moments? According to Rousseau, by the act of social union a moral and collective person, endowed with the general will, is thereby brought into existence and thereafter known as the State or Sovereign. Nor, he is careful to add, must this person be regarded as fictitious because not a man.[7] For Bosanquet, the State and its system of law and order represents my "higher" self, and its actions, even those which I explicitly reject, ought to be accepted because they are willed by the General Will which is my "real" will as opposed to my selfish and trivial actual will. The earlier quotation from Burke expresses a similar view. The essential point of this view is that the State or Society (no distinction is usually made between them) is something of far greater value than any or all of the individuals at any one time who compose it. What it ordains, therefore, and expresses as law must be good and must, therefore, be what I should also will if I were as wise as it is. In fact, it is what my higher self wills though I do not. Therefore, I "really" will it and I ought to do what I "really" want to do although I actually don't want to do it. I do not intend to examine all the linguistic shifts and ambiguities of this theory. The use of "self" and "will," e.g., and the tendentious use of "higher" and "lower" where difference of value should be proved and not merely asserted. In fact, the extremely perverse use of language by these philosophers often blinds one to the undoubted facts which they emphasise and which are neglected by alternative theories. The State is not identified with any one or with the whole collection of its members but is something over and above them. That is to say, propositions can be made about the State which would be nonsense if made about any or all of its members. E.g., "The English State has been established for at least four hundred years." The State is a moral person. That is to say, it is sensible to say of actions which we ascribe to the State that they are right – or wrong. The sense in which these words are used is different from that for individuals but

7. The Social Contract, Chapters 6 and 7.

it is not nonsense to say, "I think the State acts rightly in providing Old Age pensions," though the analysis of this statement would be very complicated. The State is greater and more permanent than the individual. That is true. The generations pass, but English state power remains. It is, therefore, likely that laws made according to the Constitution will be such as may reasonably be accepted. There is a presumption in their favour. Nor will responsible citizens wish rashly to destroy an established power and order which has served the past and serves the present moderately well and may be valued by future generations. The State serves more important purposes than the individual. Without accepting a mystical march of history, this also may be admitted. The actions of few individuals, e.g., are likely to affect so many people for so long a time as do most State actions. The State has international functions, relations to colonies, to other States which may have lastingly bad or good effects. No individuals, in their private capacity, could perform such functions. None of these facts about the difference between the State and the individual need be denied. Indeed, they are important criteria for the use of "political obligation." But they are *differences* merely. It does not follow from them that the State is either morally better or worse than any individual. Nor are any or all of them a sufficient basis for political obligation. I do not mean by saying that "I ought to obey this law," that I was born, without my choice, a member of English society from which I received education, culture, the means of livelihood and the general system of law and order without which these would be impossible. Even though it may be true that all that one is and can do, is due to the facilities provided by the State and the social order, it does not follow that all State action is right and that all laws should be obeyed. For it is not self-contradictory to say "This is an English law but it is a bad law and ought not to be kept." Moreover, this view leads to the absurd conclusion that the laws of any community are equally good. The Nuremberg laws, therefore, are good for the Germans, though they are bad for everyone else. But if they are bad, they are bad also for the Germans, though they may not recognise this.

Then is it perhaps because of their social effects that laws ought to be obeyed? Do we mean by the State, the dispenser of social benefits on the largest possible scale? This is the Utilitarian or realist view of the State. The State, like any other institution, is justified by its works. It does not depend on mythical contracts or mystical organisms but on a pragmatic sanction. "The State is an organisation for enabling the mass of men to realise social good on the largest possible scale" and social good "consists in the unity our nature attains when the working of our impulses results in a satisfied activity."[8] I am

8.   Laski. *Grammar of Politics*, pp. 24, 25. London, 1925.

justified in obeying the laws only if I am tolerably satisfied with my life in the community. And anyone's life will or ought to be tolerably satisfactory if his "impulses" are being satisfied. But what is the criterion that they are? I can know whether I am satisfied with my life, but even if I am, what follows? Does it follow that I am justified in obeying the laws because the State does or will provide the conditions of such satisfaction? Or that I am entitled to rebel if my impulses are not satisfied? Could I not be satisfied as the result of a bad law and dissatisfied as the result of a good? But Laski would say, it will be urged, that the good must be social. State action should be approved only if it promotes this desirable unity for all or most people. But what exactly is this desirable state and how do we know whether it has been achieved for everyone? Without an adequate criterion of this, how do we know what laws to obey and what governments to approve? The conditions of personal satisfaction are numerous and many could not be provided by the most benevolent State. I do not wish to suggest that Laski, or any other philosopher, supposes that all the conditions of a happy life can be provided by the State. But it is not easy to see from his remarks on the social good why any should be excluded. The utilitarian criterion, which seems so practical, is not one of the easiest to apply, or even to state clearly. Bentham's criterion is at least clear, if impossible to accept. Once this is discarded what is the "social good" or the "general welfare" whose promotion is the purpose of the State and the criterion of the goodness or badness of its actions and laws? Only the vaguest statements ever seem to be offered. This may be condemned as pedantic. To say that the State is justified by its works is to say, and this is known by all but the perverse, that it should be judged by the way in which it makes possible for all citizens the material and cultural conditions of good living. The laws of a State should be chiefly directed to securing for all its members, employment, a reasonable income, health, education, good houses, etc. Certainly not all the conditions can be provided communally but a great many can and should, and the more a State provides the more it should be approved. In fact, only if it tends to maximize certain obvious benefits should any law be obeyed. No laws ought to be partial. That is, I think, the point of the theory, which is generally favoured by social reformers. Far be it from me to minimise its importance. But again, I think, it is not and cannot be the whole story. For is it not conceivable that all these desirable objects might be promoted by what everyone would call a bad, e.g., a completely tyrannous State? The government of such a State might be exceedingly efficient in promoting social welfare to obtain popular support and the majority of its citizens might be thoroughly satisfied that all their impulses were

satisfactorily fulfilled. Ought one then to support such a government and respect its laws? The usual reply is, "Ah, but there must always be some important impulses left unsatisfied by such a Government; no bad governments could possibly promote the general good." But now, is this an empirical statement? Suppose people no longer feel these important impulses, why should they bother about their satisfaction? Does this remove the difficulty? If people no longer resent actions which would normally be called tyrannical, do they cease to be tyrannical? Is contented slavery not repulsive? Utilitarian philosophers would probably not agree. They would say, as most people would, that such actions *ought* to be resented, and that they are not does not make them good. But then the utilitarian view can surely be expressed in the tautology, "Only the governments which we ought to support are those which we ought to support because they promote general social good." This is not very enlightening.

The utilitarian view, then, which pictures the State as an institution or association for promoting the interests of its members is not adequate. The picture likens the State to any other association with a specific purpose, e.g., a trade union, a college, a commercial company or a church. But the difference is that the objects of these associations can all be fairly clearly stated and, indeed, must be before they are given legal status. The object of the State itself cannot be thus stated. High wages and good working conditions might conceivably be achieved by other means than combining in trade unions: a copper mine might be discovered and worked without floating a commercial company, but there is no describable purpose or object of social life as such any more than of human life as such, which could be obtained by some alternative means. This picture then remains as inapplicable as the others. But it, too, points to important criteria for our use of "political obligation." That some laws promote desirable social improvements in the general conditions of living for the majority of people, is a good reason for accepting them. But it is not the sole justification for accepting any and every law.

What then is the answer to the original puzzle, "Why should I obey *any* law or acknowledge the authority of *any* State or Government?" to solve which these pictures were invented. The discussion of the three most prominent types of answer seems to show that even discounting nonsense or picturesque terminology, none of them alone is sufficient. May it not also suggest that no such general answer is either possible or necessary? This would not be surprising. A general proof of the existence of material objects seems impossible, and to ask for it, absurd. No general criterion of all right actions can be supplied. Similarly, the answer to "Why should I obey

any law, acknowledge the authority of *any* State or support *any* Government?" is that this is a senseless question. Therefore, any attempted reply to it is bound to be senseless, though it may perform certain other useful or harmful functions. It makes sense to ask "Why should I obey the Conscription Act?" or "Why should I oppose the present German Government?" because by considering the particular circumstances and the characteristics of all concerned, it is possible to decide for or against obedience and support. We all know the kind of criteria according to which we should decide these two issues. But although it looks harmless and even very philosophical to generalise from these instances to "Why should I obey *any* law or support *any* government?" the significance of the question then evaporates. For the general question suggests an equally general answer and this is what every political philosopher has tried to give. But no general criterion applies to every instance. To ask why I should obey *any* laws is to ask whether there might be a political society without political obligations, which is absurd. For we mean by political society, groups of people organised according to rules enforced by some of their number. A state of anarchy is just not a state of political society and to ask whether, since laws are not obeyed in the first state they ought to be obeyed in the second is to ask a nonsensical question. But neither does it follow, as some idealists seem to suppose, that all laws should be equally accepted because commanded by a political authority. For this is, in fact, only another attempt to find a general criterion for political obligation. But it is not that which we always apply when considering political action. The political theorists want an answer which is always and infallibly right, just as the epistemologists want a guarantee that there are material objects or that generalisation to the unexamined must be valid. But these are all equally senseless requests, for they result from stretching language beyond the bounds of significance. I know how to determine on any particular occasion whether or not I am suffering from an optical illusion. Therefore, it is sensible to ask "Is this line really crooked or does it only seem to be?" But to ask whether, after applying all the relevant tests unsuccessfully, I am still and always deluded, is senseless. For the word "deluded" has now lost all significance since however hard and carefully I look I can never find a veridical perception with which to compare my delusions. The word "delusion" no longer significantly opposed to "veridical" becomes meaningless. Similarly, I can determine whether or not I ought to observe the Education Acts or the Income Tax law. Obviously, I think I ought partly because they were passed by a freely elected Parliament, according to all the usual procedure, so that in some complicated and indirect sense I have consented

to them. Then, too, they promote useful social ends and there may be other criteria for rightly obeying them. One or two of these criteria might be absent and I should not think it right to resist, if too many of them were I might get restive but not yet rebellious. A trade unionist, e.g., might rightly think that the Trade Union Disputes Act passed after the general strike was harsh and unfair but not sufficient in itself to risk civil war about, especially since a new government which trade unionists could help to elect might repeal the Act. But if too many acts are passed in suspicious circumstances and with dubious objects, the duty to resist tyranny will overrule the duty to obey law. When or how cannot be stated in advance. Nor can the criteria for accepting a law be precisely stated. Consent, tradition, objects promoted, all the criteria emphasised by the political theorists are important, but not all are equally important on every occasion, though if one or more were persistently absent over a long period we should, rightly, object. The manner in which they (and probably others) are blended is indefinitely various and no precise definition could describe our usage. Nevertheless, it does not in the least follow that we do not very often know that a law should be obeyed and a government supported and sometimes that both should be resisted. Just as we know very well that the pillar box is red and that Jane Austen was not vulgar, although both "red" and "vulgar" are used vaguely.

This may seem a disappointing conclusion and not likely to have the stirring effects of the homeric stories of the social contract and the "higher" self or even of Burke's rhapsodies on the British Constitution. But I think it has some practical value. The general, metaphysical theories are really very simple. They seek to reduce all political obligation to the application of an almost magical formula. All laws which should be obeyed result from the social contract, or the general will, or promote the greatest happiness of the greatest number, so in order to know your political duties look for the trade mark and leave the rest to government. They do imply that we can know once and for all almost by learning a single sentence, how and when political obedience is justified. But if there is no general criterion, but an indefinite set of vaguely shifting criteria, differing for different times and circumstances, then it may often, if not nearly always, be necessary to scrutinize our political relations to see whether we are on this particular occasion justified in giving or withholding our support to a measure or a government. The value of the political theorists, however, is not in the general information they give about the basis of political obligation but in their skill in emphasizing at a critical moment a criterion which is tending to be overlooked or denied. The common sense of Locke and the eloquence of Rousseau reinforced and guided

the revolt against dogmatic authority by vividly isolating and underlining with the contract metaphor the fact that no one is obliged to obey laws concerning none of which he has had a chance to express consent or dissent. It does not follow that this is the sole criterion of political obedience, still less that having derived all political obligations from a social contract or a general will we can accept them all happily and go to sleep. As rational and responsible citizens we can never hope to know once and for all what our political duties are. And so we can never go to sleep.

'NEVER GO TO SLEEP': POLITICAL THEORY AS A FOREIGN LANGUAGE

Commentary by Lorna Finlayson

LORNA FINLAYSON studied
for both her BA and PhD
in Philosophy at King's
College, Cambridge, where
she was then appointed to
a four-year junior research
fellowship, also acting as
an affiliated lecturer at
the Faculty of Philosophy.
She is the author of *The
Political is Political*: *
Conformity and the Illusion
of Dissent in Contemporary
Political Philosophy*
(Rowman & Littlefield,
2015), and *An Introduction
to Feminism* (Cambridge
University Press,
forthcoming November
2015). She has interests
in political philosophy
and its methodology,
critical theory and theories
of ideology, feminist
philosophy, philosophy
of social science, and
Arabic philosophy. As of
September 2015, she is
Lecturer in Philosophy at
the School of Philosophy
and Art History at the
University of Essex.

'NEVER GO TO SLEEP': POLITICAL THEORY AS A FOREIGN LANGUAGE

Commentary by Lorna Finlayson

I KNOW LITTLE ABOUT either Margaret MacDonald or the philosophy of language, but having recently been forced to read her article 'The Language of Political Theory', I have found myself puzzled and intrigued.

Was that an unusual beginning for an academic comment piece? I would suggest that it was, but I've modelled it on MacDonald's own opening lines. 'I have rashly chosen a subject about which I am more than usually likely to talk nonsense,' she writes, in tone which is neither simply blunt nor merely affected. 'Arch' is perhaps the word: MacDonald's words have a carefully studied, knowing air – and an undeniable touch of affectation; but at the same time they project not vanity or frivolity but a seriousness which becomes unmistakeable in the pages that follow. In saying that she knows little about history or politics, she does not seem to be apologising, nor expressing deference to the 'experts'. As she continues, she presents herself as a fascinated observer, looking on with an eye that is at once critical and cautious at the what to her is in some ways alien territory:

> '...having recently been forced to read a fair amount of what is called political theory or political philosophy, I have become both puzzled and interested by the curious notation in which much of it seems to be written. One meets here a "contract" which one is carefully warned was never contracted; an "organism" unknown to biology; a superior "person" or higher "self" with whom one can never converse; an "association" or "corporation," whose objects are obscure and which is not listed in any of the recognised Directories.' (MacDonald 1941: 108)

So the first thing that struck me about MacDonald's article was its conscious adoption of the position of the *outsider*. And the first thing that I would note about this position is that it is one which rarely finds a place in the increasingly specialised discipline of contemporary philosophy (including political theory). In order to offer a contribution to this or any other branch of the discipline now, it is increasingly expected that you will have, or will have to feign, a certain kind of expertise. If you want to write something about political philosophy (at least, if you want to have it published), far from beginning with an admission of relative ignorance, one of your obligations is to demonstrate your knowledge of

'the literature'. In-depth familiarity with a field is an entry requirement. Often, it is also necessary to show that you are up to speed with the very latest developments in the debate around your chosen topic – any shortfall in these respects equals a failure to 'engage'. You must pay your dues before you are entitled to pronounce.

One of the costs of this state of affairs, in my view, is that it erects a barrier to the kind of insight that outsider perspectives can provide – for sometimes, things may be seen clearly from the outside that cannot be seen from within. It also creates a strong imperative to work within fashionable frameworks, rather than either forging new ones or revisiting those that may have been neglected. It is indicative of a notion of 'expertise' – the experienced insiders, the professionals, know best – which is not obviously appropriate in the context of political discussions, since the very issues we might want to talk about have such bearing on the question of who (if anybody) counts as an expert. It affirms a disciplinary and sub-disciplinary 'division of labour' – political philosophers work on these questions, philosophers of logic and language on these others – which only really makes sense if you assume that each specialist camp knows what it's doing and can be trusted. It may take an outsider voice to say not only that some are playing the game badly, but that the game itself is the wrong one. Yet outsider voices are, by definition, precisely those who are not entitled to judge.

MacDonald is not trying to trash political theory as it was practised in her context (the context of Britain – and in particular, Oxbridge – in the 1930s and 1940s). Trashing is not the only or, in many cases, the best thing that an outsider may do. Instead, she is asking questions about how best to make sense of the things that political theorists say, their distinctive sorts of claim and their stock disputes. It is harder to ask those questions from the perspective of a political theorist. Or perhaps, the point is that it is easy not to, if you are someone who is actively and contentedly embroiled in making and arguing about these claims and entering into these disputes.

MacDonald's status as a non-specialist affords her the freedom and critical distance which make her piece such an interesting one. I cannot discuss anywhere near everything that I found of interest in it. It is an unusually rich paper, which (again in sharp contrast with the style of academic philosophy currently dominant) does not proceed by identifying one self-contained 'problem' to be solved or a single determinate thesis to be established, but has instead the sense that the author is looking around at her unfamiliar surroundings with wide-open eyes and noting things of interest.

To today's reader, MacDonald's semi-affected bewilderment at constructions such as the hypothetical contract or the 'higher self' will be immediately evocative of contemporary 'realist' critiques of the style of political theorising which has become mainstream – a style dominated by various forms of abstraction and 'ideal theory'. According to realists, political theory should instead begin from the concrete realities of history and politics. But it would be a mistake to assimilate what MacDonald is saying, without remainder, to 'realism', at least in its contemporary guise. She does not seem to be saying that these constructions, or the points that are made on the back of them, are always false or inappropriate – although she certainly strikes a note of scepticism when she observes that devices like the hypothetical social contract have very little to do with the decisions and practical problems we actually face: 'Nor is your depression at the Labour Exchange likely to be much relieved by being told that you "really" willed your unemployment (you would never have thought so, unaided) or that the State is a very superior moral person, only even more anonymous and inaccessible than the Permanent Secretary to the Ministry of Labour.' (MacDonald 1941: 109) She is less concerned to tell political theorists what to do than to try to make sense of what they are doing (and what they take themselves to be doing). What her wryness of tone might be taken to express is just that there is something peculiar going on here – what is being invoked is nothing detectable by the senses, nor an explicitly moral or linguistic claim – and that there is something especially strange about offering this sort of story in response to what are fundamentally *practical* problems. The strangeness needs emphasising because it is such a *familiar* strangeness – which is why it takes a stranger to see it.[1] This does not necessarily mean that political theorists are doing anything wrong or stupid, but it does at least seem to mean that we are owed an account of what is going on.

It's also important to bear in mind the differences between the context in which MacDonald was writing and the context of contemporary academic philosophy. She is not preaching to political theorists on the need to take history and real politics seriously – not only because, professing to know little about either, she is not in the strongest position to do so, but also because the context of the 1930s and 1940s was not yet one in which it was normal for political theorising to be as proudly ahistorical as it often is today. Nor were disciplinary boundaries so prominent, rigid, and aggressively policed as now. The use of abstraction – and of particular abstractions such as the 'higher self' (more likely now to be couched in terms of 'higher-order desires', or in terms of the 'idealised observer') – is common to both contexts.

1. Similarly, I notice that my students quickly learn to use the phrase 'conception of the good', as if it were perfectly normal. Anyone not initiated into the vocabulary of (post-) Rawlsian political philosophy would find this conept very strange, and rightly so. Not even political philosophers – and perhaps not even Rawlsians – actually seem to think of themselves as having and refining these conceptions, and most would likely be startled by the question, "So, what's *your* conception of the good?".

But MacDonald is too sophisticated to think that idealisation or abstraction per se is either good or bad – it all depends on the *kind* of abstraction, its manner and function: 'In political theory, as, indeed, in philosophy generally, it is very often not what is said, but the spirit in which it is said, which makes the difference.' (MacDonald 1941: 113-14)

The context in which MacDonald wrote this particular piece, at the beginning of the 1940s, was one in which logical positivism was the most salient feature of the landscape of British academic philosophy (the Hegelian-inspired approach of the 'British idealists' was by the 1930s largely out of favour). What would soon come to be known as 'ordinary language philosophy' was nascent in the work of the later Wittgenstein (who had been a teacher and mentor of MacDonald's during her time as research fellow at Girton College, Cambridge). These are the currents which inform her response to the peculiar abstractions of political theory (and when she approaches these, she is of course *not* approaching the barren desert of post-Rawlsian liberal thought which confronts contemporary realists).

Although ordinary language philosophy developed into an autonomous school with its distinctive methodology (basically: find out how ordinary people use words), both this strand and the tradition of logical positivism share the conviction that philosophical questions may be solved through linguistic analysis. For logical positivists, linguistic constructions like claims or questions always fell into one of three categories: claims or questions about empirical states of affairs; claims or questions about the use or meaning of words; and literal nonsense (statements or questions that say or ask nothing at all). Political theory was no exception. The task was to sort its pronouncements into these three categories – and the result was not usually pleasing to political philosophers.

What MacDonald argues is that much of what is said in the context of political theory falls into none of the above three categories: it has no empirical content; it is not plausibly (only) about the use of language; and yet it is not nonsense – or if it is, it is *important* nonsense. What she is saying here is not only that nonsense may have effects (serious or otherwise) – something which nobody was really trying to deny, after all. MacDonald's more interesting claim is that these effects are part of what she calls the 'philosophical "point"' of what is said. She makes the importance of the distinction very clear in the following passage:

'No one will deny that in political affairs philosophical nonsense may have serious effects. Is this philo-

sophically relevant, or not? To deny that it is seems to reduce philosophising to mere scholastic verbalism. Not for any moral reason. Not because philosophers ruin themselves and their subject by sitting in ivory towers and talking about the uses of words instead of considering how the people perish (or flourish) on nonsensical theories and slogans. But simply because, not to try to understand how this language has effects, even though it may give no information, is to miss half its philosophical point and so is bad philosophy. The philosophical "point" of a remark (or the "point" of a remark which is of philosophical interest) is, at least partly, connected with the cause or reason which induces people to go on making it, though it can neither be supported nor refuted by any empirical evidence. It may be false, it may, if taken literally, be meaningless, but they feel that it has some use. This does seem to be relevant to the understanding of some philosophical remarks, if not of all.' (MacDonald 1941: 110)

What is being said here, as I understand it, is that you cannot (always) separate the issue of what effects something has from the question of what sort of thing it is and what it is for (its 'point') – and the effects it has will be connected in turn to the kind of context out of which it arises ('the cause or reason which induces people to go on making it'). Whilst for MacDonald there is a sense in which some statements including many in political theory may be meaningless – that is, they may neither carry any empirical content nor be plausibly reducible to linguistic recommendations – there is another crucial sense in which they *do* have meaning, and this meaning (or 'point') is to be found not in any proposition conveyed but in the source and function of what is said. It is not *always* appropriate, or not always equally so, to connect the meaning or point of something to its effects – and this is where the importance of the *source* or *origin* comes in. The solipsist does not refuse to converse with others ('unless he is also suffering from incipient schizophrenia'), and that reflects the fact that the 'problems' to which solipsism is offered as a response are not *practical* problems (they 'have not, usually, been suggested by any practical difficulties about communication or knowledge in ordinary life'). (MacDonald 1941: 111) The case is not the same for claims about ethics or politics, argues MacDonald. The latter arise out of real and often pressing human problems, and have real human effects. The conclusion she ultimately draws from this is not a verdict for or against any particular thesis as to the proper relationship between the state and the citizen – such as Rousseau's doctrine of the 'general will' or Burke's defence of the authority of the state – nor any other verdict about how things should be, politically speaking. The contribution of political theorists – and it can be a genuinely useful and important one – is to highlight particular aspects of the world, to make possible (through the use of a kind of poetic imagery) par-

ticular ways of looking at the world at particular times. What we cannot do is lay down a theory of 'political obligation' which captures, once and for all, the truth about where our political obligations come from and what they are – so that we 'can accept them all happily and go to sleep.' (MacDonald 1941: 124) This cannot be done because the statements of political theory were never in the business of capturing or failing to capture truths about the world. They are, in the last instance, practical interventions in response to practical situations – and practical situations, as we know, change all the time. MacDonald's conclusion: 'As rational and responsible citizens we can never hope to know once and for all what our political duties are. And so we can never go to sleep.' (ibid)

One of the first things that I was taught about ethics as an undergraduate was that the 'ethical' and the 'metaethical' were autonomous realms: just as you couldn't (i.e. were not allowed to) infer an 'ought' from an 'is', it was fallacious to infer any 'normative' conclusion (about what should be done) from a metaethical premise (such as an attachment to moral 'relativism').[2] This never seemed quite right to me, but it was made pretty clear to us that it was not something that was up for legitimate question (to question it just meant that you still hadn't understood something quite basic). As MacDonald observes, however, to act differently on the basis of some conviction about the nature of ethics doesn't *seem* crazy, in the way that it would seem crazy for the solipsist to refuse to talk to anybody, or for the sceptic about the external world to 'sit down very gingerly on every chair for fear it isn't really there.' (MacDonald 1941: 111) And this is still so when the thesis in question carries no empirical content, and even when it seems to belong to the 'meta-' level of reflection on ethics rather than to the domain of 'normative' prescription. MacDonald gives the example of the claim that 'all action is "really" selfish', and judges that to say this 'is to make not an empirical but a grammatical statement'. As such, a statement like this tells us nothing about what to do (other than that we should use words in a certain way). And yet, she observes, 'many people would not think it at all absurd that anyone who said, "An action is right only if by doing it the agent will promote his own advantage" should so interpret this as to neglect most of the actions that would ordinarily be called his duties.' (ibid.)

Of course, those who hold to the view I was taught as an undergraduate will say that the mere fact that people commit a fallacy does not stop it from being a fallacy, and that what 'seems' right to us may not always be so. Fine. But what MacDonald is doing is asking us to reflect on the asymmetry between the way we re-

2. Cf. Bernard Williams's critique of 'vulgar relativism' in Williams 1972.

spond to different kinds of claim: claims with the same kind of structure or status (e.g. nonsensical, or apparently analysable only as linguistic recommendations), but dealing with different areas of life and arising from different kinds of problem. *Why* are we often inclined to see metaethical claims as having a bearing on the way we act, when we do not react the same way to claims about evil demons? The answer that MacDonald is suggesting, as I read it, is that *we're not wrong* (or not wrong in every sense, anyway). Rather, it is a certain kind of thinker who is wrong, because the thinker who cannot see (or is not interested in) the effects of our ways of speaking, and how these effects relate to the problems to which these ways of speaking are a response, has missed 'half [their] philosophical point'. MacDonald has something important to say to contemporary philosophers, if only because so many of us are taught to be precisely that kind of thinker.

REFERENCES

MacDonald, M.. (1941), 'The Language of Political Theory', *Proceedings of the Aristotelian Society* 41: 91-112.

Williams, B. (1972), *Morality: An Introduction to Ethics*. New York: Harper & Row.

SOME REFLECTIONS ON MORAL-SENSE THEORIES IN ETHICS

by C. D. Broad

CHARLIE DUNBAR BROAD
(1897 – 1971) was born
in London and educated
at Trinity College,
Cambridge, initially as
a natural scientist. He
shifted to philosophy, and
held posts successively at
Cambridge, St Andrews
and Bristol before returning
to Cambridge in 1923.
He became Knightbridge
Professor of Moral
Philosophy there in 1933
and held the position until
his retirement twenty years
later, during which time
found himself increasingly
isolated in a philosophical
environment dominated
by the influence of Ludwig
Wittgenstein and his
disciples. He worked on
various topics in ethics,
philosophy of mind, and
metaphysics. His most
significant works were
*The Mind and its Place in
Nature* (London: Kegan
Paul, 1925), *Five Types of
Ethical Theory* (London:
Kegan Paul, 1930), and his
magisterial *Examination
of McTaggart's Philosophy*
(Cambridge: CUP, 1933,
1938).

SOME REFLECTIONS ON MORAL-SENSE THEORIES IN ETHICS

by C.D. Broad

DURING THE LONG VACATION OF 1944 I spent such time as I could spare from my other duties in reading with some care Richard Price's book *A Review of the Principal Questions and Difficulties in Morale*. This was first published in 1758, and it reached a third edition in 1787. Price died soon afterwards, viz., in 1791. Until Ross published his book *The Right and the Good* in 1930 there existed, so far as I know, no statement and defence of what may be called the "rationalistic" type of ethical theory comparable in merit to Price's. Price was thoroughly well acquainted with the works of other great English philosophers and moralists, such as Locke, Berkeley, Hume and Butler, and he develops his own views in conscious opposition to those of Hutcheson, the founder of the so-called "moral-sense" type of ethical theory.

I had thought at one time of writing a critical account of Price's doctrines. But, when I began to do so, I soon found that it would be more profitable to treat independently and in modern terminology some of the questions with which Price was mainly concerned. Therefore my further references to Price will be only occasional and incidental; but I wish to make it plain that his book is the background of my paper, and that reading the former was the stimulus to writing the latter.

The topic with which I shall be primarily concerned may be called the "epistemology of moral judgments." This subject is of considerable interest in itself, and I think that it has been very inadequately treated by most writers on ethics. But it is important also for another reason. Questions of epistemology and of logical analysis are interconnected, and the answer which we give to a question of the one kind may have an important bearing on that which we should be inclined to give to a question of the other kind, e.g., I should be prepared to argue that, if ethical terms, such as *right* and *good*, are simple and non-naturalistic or are complex and contain a non-naturalistic constituent, then the concepts of them must be wholly or partly *a priori*. On the same hypothesis I should be prepared to argue that such judgments as "Any act of promise-keeping tends as such to be right" must be synthetic and *a priori*. Now it is a well-known and plausible epistemological theory that there are no *a*

priori concepts and no synthetic *a priori* judgments. If I am right, anyone who feels no doubt about this epistemological theory can safely reject the analysis of moral judgments which makes them contain non-naturalistic constituents. On the other hand, anyone who feels bound to accept that analysis of moral judgments will have to reject this epistemological theory.

In the discussion which follows I shall confine myself to the concepts *right* and *wrong*, in the specifically moral sense, and to judgments in which they occur as predicates. I think that most of what I say could be transferred *mutatis mutandis* to the concepts *morally good* and *evil* and to judgments in which they occur.

As Price points out, the words "right" and "wrong" are used in at least two different senses. This is made obvious by the fact that the sentence "It is always right for a person to do what he honestly believes to be right, and wrong for him to do what he honestly believes to be wrong," is intelligible and would generally be admitted to be in some sense true. The two senses in which "right" and "wrong" occur in this sentence may be described as the "subjective" and the "material." An act is subjectively right if the consequences which the agent *expects* it to have are such as he thinks would be materially right in the situation as he *believes* it to be. We shall be concerned here only with material rightness and wrongness. Let us call sentences in which the words "right" or "wrong," used in the material sense, occur as predicates "deontic sentences." An example would be "Any act of promise-breaking tends as such to be wrong."

I shall first distinguish certain alternative analyses which have been proposed for the situations expressed by deontic sentences, and then I shall consider certain alternative theories which might be held concerning deontic knowledge or belief. In the course of the discussion I shall try to bring out the relations between the two sets of theories.

1. ALTERNATIVE ANALYSES OF DEONTIC SENTENCES

When a person utters such a sentence as "That act is right" he seems *prima facie* to be expressing a judgment, and in that judgment he seems *prima facie* to be ascribing to a subject a predicate which has no reference to his own or other men's sensations, emotions, desires, or opinions. But we know that such appearances may be misleading. Such sentences as "This food is nice" and "That thing is yellow" are of the same grammatical form as "That act is right." Yet everyone would hold that the predicate of the first refers to the speaker's sensations of taste, and many people would hold that the predicate

of the second refers to the visual sensations of human beings. So the first question to be asked is this. Do sentences like "That act is right" express judgments at all? If not, what do they express?

As is well known, there is a theory that such sentences do not really express judgments at all. It has been held that they express only certain emotions felt by the speaker, or certain desires of his, or certain commands. I shall call this the "Interjectional Theory." Price does not consider this extreme view. If it had been put to him, he would probably have regarded it as too fantastically absurd to be taken seriously. It is, indeed, the kind of theory which can be swallowed only after one has undergone a long and elaborate process of "conditioning" which was not available in the eighteenth century.

Suppose that the Interjectional Theory is rejected. Suppose we hold that deontic sentences do express judgments of *some* kind, and that at any rate the fact that they are in the indicative mood is not misleading. The next suggestion is that the judgments which they express are really about certain human experiences, certain sensations or emotions or desires. I shall call this the "Subjective Theory." I shall now point out that it may take a great number of different forms, and shall try to classify them.

The factor common to all forms of the Subjective Theory is that there is a peculiar kind of experience which human beings are liable to have when they contemplate certain acts, e.g., acts of promise-keeping or of treachery, just as there is a peculiar kind of experience which they have when they look at certain objects, e.g., at snow or at soot. I propose to call this at present by the intentionally vague name "moral feeling." I use this term because it covers both sensation and emotion. Since deontic judgments take the two opposite forms "That is right" and "That is wrong," it must be assumed that moral feeling takes two opposite forms. There are analogies to this both in sensation and emotion. There are the opposed temperature-sensations of hotness and coldness, and there are the opposed non-moral emotions of love and hate. I shall speak of the "pro-form" and the "anti-form" of moral feeling, and will assume that the former is associated with judgments of rightness and the latter with those of wrongness. The first division of Subjective Theories is into *Sensational* and *Emotional*, according to whether moral feeling is held to be analogous to sensation and moral judgment to be analogous to judgments of sense-perception, or whether the feeling is held to be a form of emotion and the judgments to be concerned with that emotion.

The next division of Subjective Theories is into what

I will call the "Intra-subjective" and the "Trans-subjective" varieties. According to the first of these a person who judges that so-and-so is right is asserting something about *his own* moral pro-feelings only. He is not saying anything about the moral feelings of other men. According to the second variety such a person is asserting something about all men, or most men, or a certain restricted class of men, and not only about himself.

Lastly, each of these two varieties of the Subjective Theory can be subdivided into what I call an "Occurrent" and a "Dispositional" form. On the occurrent form of the intra-subjective variety of the subjective theory a person who says that so-and-so is right is asserting only that at this moment he is having a moral pro-emotion towards so-and-so. On the occurrent form of the trans-subjective variety of the theory he is asserting that all or most members of a certain class of men, e.g., most members of the Athenaeum, are at present having a moral pro-emotion towards so-and-so. On the dispositional form of the intra-subjective variety of the theory he is asserting that he has a disposition to feel a moral pro-emotion whenever he contemplates so-and-so or other acts like it. He may not be feeling such an emotion at the moment when he is saying that so-and-so is right. He might not be actually witnessing or thinking of such an act at the time; or, if he were, he might be in some special occurrent state, such as anger or jealousy, which is inhibiting or reversing his disposition to feel moral pro-emotion. On the dispositional form of the trans-subjective variety of the theory he is asserting that all or most men or all or most members of a certain class of men have a disposition to feel moral pro-emotion when they contemplate so-and-so or other acts like it. He might have strong reason to believe this even if he lacked that disposition himself. I have, e.g., strong reason to believe that most men have a disposition to like the taste and smell of apples, though I personally loathe them.

It appears then that there are at least eight possible species of the Subjective Theory, according as it is (i) sensational or emotional, (ii) intra-subjective or trans-subjective, or (iii) occurrent or dispositional. There are two remarks that I would make at this point.

(i) Even on the occurrent intra-subjective form of the theory such a statement as "That act is wrong" could be questioned without accusing the speaker of lying about his own feelings at the time. But this could happen only in one way. The speaker might be mistaken about the kind of feeling which he is having when contemplating this act. He might think that he is having a moral anti-feeling when really he is having what Sidgwick calls a "feeling of quasi-moral repugnance." I have no doubt

that such mistakes are often made by people, e.g., about their own feelings towards abnormal sexual desires and practices. (ii) I am inclined to think that the only form of the theory that is worth serious consideration is the trans-subjective dispositional form of it. But I should admit that it is not implausible to hold that *sometimes* when a person says that so-and-so is right or that it is wrong he may be talking only of his own disposition to have a moral pro-feeling or anti-feeling when he contemplates such acts.

So far I have spoken only of singular deontic judgments, i.e., those of the form "That act is right (or is wrong)." But there are also universal deontic judgments, such as "Any act of promise-keeping tends as such to be right" or "Any act of deliberate deception tends as such to be wrong." How would the Subjective Theory deal with the latter? Let us take, e.g., the trans-subjective dispositional form of the subjective theory and consider how it would deal with "Any act of promise-keeping tends as such to be right." It would say that this is equivalent to "Any person contemplating an act which he believed to be one of promise-keeping would tend to have a moral pro-feeling in so far as he confined his attention to that aspect of the act." No doubt this might require various qualifications, e.g., we might have to substitute "any *normal* person" for "any person" in order to allow for moral lunatics, and we might have to add "provided he were in a normal state at the time" in order to allow for the possibility of his disposition to have a moral pro-feeling being inhibited or reversed if he were in a state of rage or of jealousy. But the general principle is clear enough.

Next let us suppose that all forms of the Subjective Theory are dismissed. We should then have to accept some form of what I will call the "Objective Theory." According to this a deontic judgment ascribes to an act a certain quality or relation or relational property which has no reference to the feelings or desires or opinions of the speaker or of anyone else concerning that act. Such judgments would be significant and might be true even if no human being had ever had moral feelings of any kind.

No doubt the Objective Theory might take many different forms. But for our purpose the most important principle of division is the following. Let us describe an "ethical sentence" by enumeration as any sentence in which the words "right" or "wrong," "ought" or "ought not," "morally good" or "morally evil" or mere dictionary equivalents of them occur. Now, if the words "right" and "wrong" denote opposite forms of a certain objective characteristic, the following possibilities are

open about that characteristics. (i) It may be simple and therefore indefinable, as, e.g., the quality of sensible yellowness and the relation of temporal precedence are. (ii) It may be complex and therefore definable. If so, it may be definable (a) only by means of ethical sentences or (b) without the use of such sentences. The following alleged definitions of "right" would illustrate these two possibilities. The first would be exemplified if "right" could be defined only as "what it is *fitting* to approve" or only as "what is conducive to *morally good* experiences." The second would be exemplified if "right" were definable as "conducive to social stability" or as "productive of a balance of *pleasant* experiences." I propose to give the name "naturalistic" to (i) all forms of the Subjective Theory, and (ii) any form of the Objective Theory which holds that "right" and "wrong" are definable without the use of ethical sentences. I shall give the name "non-naturalistic" to any form of the Objective Theory which holds that "right" and "wrong" are either indefinable or definable only by means of ethical sentences. For the present purpose it is not important to consider whether this use of "naturalistic" and "non-naturalistic" agrees exactly either in extension or in intension with Professor Moore's usage.

Before leaving this topic there is one further remark to be made. I think it is fair to say that most competent persons who have reflected on this subject in recent years would agree that the only alternatives worth serious consideration are some form of either (a) the Interjectional Theory, or (b) the dispositional variety of the Subjective Theory, or (c) the non-naturalistic variety of the Objective Theory. Perhaps I should add that under the head of "competent persons" in this connexion I do not include the eminent natural scientists who from time to time take a holiday from their professional labours in order to instruct us in ethical theory.

2. ALTERNATIVE EPISTEMOLOGICAL THEORIES OF DEONTIC COGNITION

I shall begin by considering singular deontic judgments, i.e., ones of the form: "That act is right (or is wrong)." Presumably those moralists who hold a Moral Sense Theory intend at least to assert that these judgments are analogous in certain important respects to judgments of sense-perception, such as "That thing is yellow."

Now the first thing to notice is that two very different accounts may be given of such judgments as "That thing is yellow." These may be described as the Naively Realistic Account and the Dispositional Account. I will now explain these terms.

(i) I think that the plain man in his plainer moments uncritically takes for granted that the very same sensible quality of yellowness which is presented to him when he looks at a bit of gold in white light literally pervades the surface of that bit of gold, not only when he is looking at it in white light, but also and in precisely the same sense when no one is looking at it and when it is in the dark. He believes that looking at the thing and its being illuminated by white light serve only to *reveal* to him the yellowness which has been there all the time in precisely the form in which it is now presented to him. This is what I call the "Naively Realistic Interpretation." Price seems to have thought that this, or something like it, is what plain men believe. He also thought that this belief is not only mistaken, but can be seen to be internally inconsistent by anyone who reflects carefully on the natures of sensible yellowness and of material objects. I must confess that I cannot see this myself.

(ii) A person who makes the judgment "That thing is yellow" may be expressing *only* his belief that it would present a yellow appearance to any normal human being who might at any time view it in white light. No doubt a person who accepts the Naively Realistic Interpretation also believes this conditional proposition. But this belief is certainly not the whole of what he expresses by saying "That thing is yellow," and it might not even be a part of it. It might be for him only a very obvious and immediate consequence of what he expresses by that statement. I give the name "Dispositional Account" to the view that the whole meaning of such judgments as "That thing is yellow" is a conditional proposition of the kind which I have just enunciated.

The next point to notice is this. If a person believes that a certain thing would present a yellow appearance to any normal human being who should at any time view it in white light, he does not generally accept this conditional proposition as an ultimate fact. He generally amplifies it as follows. He ascribes to the thing a certain intrinsic property, and he ascribes to each human being a certain other intrinsic property correlated with the former. Let us call these respectively the "objective" and the "subjective correlate" in the perception of yellowness. It is held that when and only when a certain relationship is set up between a human being and this thing the subjective correlate in the person and the objective correlate in the thing together cause the thing to present a yellow appearance to the person.

This is common ground to the holders of the Naively Realistic and of the Dispositional Account. But there is a profound difference between them in point of detail. On the Naively Realistic Interpretation the objective cor-

relate just is that quality of sensible yellowness which, according to that theory, is spread out over the surface of the thing ready to be presented whenever the appropriate revelatory conditions are fulfilled. The subjective correlate just is the power of prehending the yellowness of yellow things when such conditions are fulfilled. That power is activated whenever a person who possesses it stands in a certain bodily and mental relation to a thing which possesses yellowness.

On the Dispositional Interpretation the objective correlate is generally held to be a certain kind of minute structure and internal agitation in a thing which is not itself literally and non-dispositionally coloured. Again, the subjective correlate is not now the power of prehending the objective correlate. We have no such power. It is the capacity to have sensations of a certain kind, called "sensations of yellowness"; and these are not prehensions of a quality of yellowness inherent in the thing perceived. There is no such quality. That power is activated whenever a person who possesses it stands in a certain bodily and mental relation to a thing which has this peculiar kind of minute structure and internal agitation.

I do not think that anyone who accepted the dispositional interpretation would give the name "yellowness" to that minute structure and internal agitation of a colourless object which, according to him, is the objective correlate of sensations of yellow. He would confine the name "yellow" to (a) the peculiar sensible quality of certain sensations, e.g., those which he has when he looks at the yolk of an egg in white light, and (b) the dispositional property which certain things have of giving rise to such sensations in a normal human observer when he views them in white light. If he were wise, he would distinguish these two usages of the word as "sensible" and "physical" yellowness; or he might prefer the more general phrases "occurrent" and "dispositional" yellowness. To the minute structure and internal agitation which are the objective correlate of the perception of things as yellow we might give the name "physical correlate of yellowness." We can now see that the Moral Sense Theory of singular deontic judgments might take two entirely different forms, viz., a naively realistic one and a dispositional one. Both would start from the common ground that there is a peculiar kind of experience which human beings are liable to have when they contemplate certain acts, and that this can take either of two opposite forms, viz., a pro-form and an anti-form. Both would hold that this experience is of the nature of feeling, where "feeling" is used to include both sensation and emotion as distinguished from thought. From this common basis they diverge as follows:

The naively realistic form of the Moral Sense Theory would take moral feeling to be like what visual sensation is supposed to be on the naively realistic view of visual perception. When a person contemplates a certain act and has a moral pro-feeling in doing so that feeling either is or involves a prehension by him of a certain characteristic, viz., rightness, in the act; and that characteristic belongs literally and non-dispositionally to the act quite independently of whether anyone happens to contemplate it or to have a moral pro- feeling when doing so. (I have used the alternative phrase "is or involves a prehension" rather than the simpler phrase "is a prehension" because it might well be held that a moral feeling is never *just* a prehension of the objective rightness or wrongness of a contemplated act, but is always such a prehension qualified by a certain kind of emotional tone.)

I am fairly certain that the adherents of the Moral Sense Theory did not interpret it in this way; for they did not, I think, put a naively realistic interpretation on visual sense-perception. But some of them may quite likely have thought that plain men mistakenly put this interpretation both on such judgments as "That act is right" and on such judgments as "That thing is yellow." On the other hand, I suspect that Professor Moore, when he compared intrinsic goodness with yellowness in *Principia Ethica*, was tacitly assuming something like the naively realistic interpretation of both such judgments.

The dispositional form of the Moral Sense Theory would take moral feeling to be either (a) a special kind of emotion or (b) a sensation analogous to those of taste or smell and not to those of sight. I suppose that hardly anyone would put a naively realistic interpretation on such perceptual judgments as "That is bitter" even if he were inclined to put such an interpretation on judgments like "That is yellow."

Starting from this basis the theory might take the dispositional form in one or other of its main varieties. The feature common to all of them would be that the moral feeling which a person has when he contemplates an act neither is nor involves a prehension by him of an independent non-dispositional characteristic of rightness inherent in that act. On the trans-subjective variety of this theory a person who says that an act is right means, roughly speaking, no more than that any normal person who should contemplate this act when he was in a normal condition would have a moral pro-feeling. On the intra-subjective variety of the theory the speaker would mean the same kind of thing with "he himself" substituted for "any normal person." I have little doubt that most upholders of the Moral Sense Theory meant to as-

sert the trans-subjective variety of the dispositional form of it. But they did not always make this clear to their readers, and perhaps they were not always clear about it themselves.

It is perhaps worth remarking that the Moral Sense Theory might conceivably take the occurrent intra-subjective form. It might allege that, when a person calls an act right, all that he means is that his present contemplation of it is accompanied by a moral pro-feeling. I think that this form of the theory is so obviously inadequate that supporters of the Moral Sense doctrine can hardly have meant to assert it. But some of them may have incautiously made statements which would suggest that this is what they meant, and their opponents may sometimes have found it convenient to seize upon these as readily assailable Aunt Sallies. It seems to me that the only two forms of the Moral Sense Theory that are worth serious consideration are the naively realistic form and the trans-subjective variety of the dispositional form. I shall now consider them in turn.

2.1. NAIVELY REALISTIC FORM OF THE MORAL SENSE THEORY

The only kinds of sense-perception which can with any plausibility be interpreted in a naively realistic way are visual and tactual perception. Therefore the naively realistic form of the Moral Sense Theory will have very little to recommend it if singular deontic judgments differ from judgments of visual and tactual perception in just those respects which make a naively realistic interpretation of the latter plausible. It seems to me that the relevant differences are profound and that the analogies are superficial.

(i) In stating the Moral Sense Theory I have so far used the intentionally vague phrase "having a moral pro-feeling or anti-feeling when one *contemplates* as action." If singular deontic judgments are to be analogous to judgments of visual or tactual sense-perception, this must be held to be analogous to having a sensation of yellowness when one *looks* at the yolk of an egg or having a sensation of coldness and hardness when one *touches* a block of ice. Is there any such analogy?

We must begin by distinguishing two cases, viz. (a) where one person makes a deontic judgment about an act done by another, and (b) where he makes such a judgment about an act done by himself.

(a) One person never can perceive the act of another, if by "act" we mean something to which moral predicates can be applied. He can perceive only some bit of overt behaviour on the part of another, e.g., writing a

cheque and handing it over to a third person. That bit of overt behaviour may be an act of forgery or of paying a debt or of subscribing to a charity or of bribing an official. As a subject of moral predicates it is a different act according to the different intentions with which it is done. Now one person can contemplate another's intentions only in the sense of making them objects of *thought* and never in that of *perceiving* them.

I think that this suffices to wreck the Moral Sense Theory in its naively realistic form as applied to singular deontic judgments made by one person about the acts of another. Even if a naively realistic account of such judgments as "That thing is yellow" were acceptable, there would be no analogy between them and such judgments as "That act is right" when the judger and the agent are different. For "that thing," e.g., a certain bit of gold, is perceived by the person who makes the judgment that it is yellow. The thing is *perceived*; it is perceived as *yellow*; and the sensation of yellowness is an *essential constituent* of the perception of the thing. The naively realistic account of the situation is that the percipient is acquainted with the surface of the thing, and that the latter reveas to the percipient through his sensation of yellow that objective non-dispositional quality of yellowness which it possesses independently of human observers and their sensations. This account is here *prima facie* highly plausible. But "that act," if done by another, is not perceived except as a bit of overt behaviour. In respect of those characteristics which make it a possible subject for moral predicates it can only be *conceived*. The moral feeling, even if it be a sensation and not an emotion only, is not an essential constituent of the perception of the act as a bit of overt behaviour; only visual sensations are essential constituents of that perception. And finally the relation of the moral pro-feeling or anti-feeling to the conception of the act as, e.g., one of debt-paying or one of bribery cannot possibly be like the relation of a sensation to a perception of which it is a constituent, e.g., the relation of a sensation of yellowness to the visual perception of a thing as yellow.

(b) When a deontic judgment is passed by a person on one of his own acts the above criticism does not hold. In performing an act a person is or may be directly aware of his own intentions. He knows it directly as an act of intended bribery or forgery or debt-paying or whatever it may be, and not merely as a bit of overt behaviour of a certain kind. Similarly, in retrospection a person generally knows by personal memory what were his intentions in his own past acts. No doubt introspective self-perception and personal memory are very different in important respects from sense-perception. But they agree with it, and differ from one's awareness of the experiences

of another person in being ostensibly instances of direct acquaintance with particulars. It seems to me then that, if the Moral Sense Theory in its naively realistic form is to be defended, it must be confined in the first instance to deontic judgments made by a person about his own acts. We might suppose that he derives his notions of rightness and wrongness from perceiving those characteristics in certain of his own acts by means of moral sensations. Once he has acquired the notions in this way he can proceed to apply them to the acts of other persons although he cannot perceive these and therefore cannot perceive their rightness or wrongness, but can have only conceptual cognition about them.

Now I think that there is a very serious objection to this view. It is certain that I have moral pro-feelings and anti-feelings *both* when I introspect or remember certain acts of my own and when I conceptually cognise the similar acts of other persons. Now I cannot detect any relevant difference between my moral feelings in the two cases. But, as we have seen, it is impossible in the latter case to hold that there is any analogy to visual sense-perception as interpreted by the naively realistic theory. It is impossible to hold here that the moral feeling is a state of acquaintance with an objective characteristic of rightness or wrongness in the cognised act. Therefore it seems unreasonable to suppose that the precisely similar moral feeling which one has when introspectively perceiving or rernembering one's own acts is susceptible of a naively realistic interpretation.

I pass now to another profound *prima facie* difference between singular deontic judgments and judgments of visual or tactual perception. If I judge that a certain act is right or that it is wrong, it is always sensible for anyone to raise the question "What *makes* it right or *makes* it wrong?" The answer that we expect to such a question is the mention of some non-ethical characteristic of the act, e.g., that it is an act of promise-keeping, of giving a false answer to a question, and so on. Let us call these "right-inclining" and "wrong-inclining" characteristics. Now the connexion between the presence of any of these non-ethical characteristics and the tendency of an act to be right or to be wrong seems to be necessary and self-evident, not causal and contingent. (I say the "*tendency* to be right or to be wrong" and not just rightness" or "wrongness" for a reason which will be familiar to all readers of Ross's ethical writings. One and the same act may be, e.g., an act of truth-telling and one of betrayal. It is not self-evident that such an act is resultantly right or resultantly wrong. But it might well be held to be self-evident that it tends to be right in respect of being an act of truth-telling and to be wrong in respect of being one of betrayal, and that it would be right

if it had no wrong-inclining characteristic and would be wrong if it had no right-inclining characteristic. These points were made clearly enough by Price, but have since been made much more clearly by Ross.)

Now the fact which I have just mentioned is relevant to both forms of the Moral Sense Theory, but for the present we are concerned only with the naively realistic form of it. If I look at a thing and judge it to be yellow, it is not particularly sensible to ask "What makes it yellow?" The question is sensible only if it is interpreted causally, e.g., in some cases the answer might be that it contains saffron. And a more ultimate answer would be that it has such and such a minute structure and internal agitation. Now on the naively realistic theory the thing is pervaded literally and non-dispositionally by an inherent quality of yellowness; and there is no self-evident necessity for all things which have a certain kind of minute structure and internal agitation and only such things to be pervaded by yellowness. It is simply a contingent general connexion between two sets of properties of a material thing, viz., certain geometrical and kinematic properties, on the one hand, and a certain objective colour, on the other. The connexion between being an act of promise-breaking and tending to be wrong does not seem to be in the least like this.

It is worth while to remark before leaving this topic that, even if our cognition of the rightness or wrongness of acts were analogous to visual or tactual perception interpreted in the naively realistic way, it is quite certain that our cognition of right-inclining and wrong-inclining characteristics is not. Such characteristics as being an intentional breach of promise, an intentional return of a borrowed article, and so one are highly complex relational properties. They can be cognised only *conceptually*; it is nonsensical to suggest that they could be cognised by anything analogous to sense-perception or to introspective self-perception.

On the other hand, the fact, if it be a fact, that the connexion between certain non-ethical characteristics and the tendency to be right is necessary and self-evident is not in itself a reason for denying that rightness and wrongness are cognised by something analogous to sensation interpreted in a naively realistic way. For the connexion between having shape and having size is necessary and self-evident, and yet both these characteristics are cognised by visual sense-perception.

I think that the upshot of this discussion is that there is little to be said for and much to be said against the Moral Sense Theory in its naively realistic form as applied to deontic judgments. We can therefore pass to the Dispositional Form of the theory.

I do not think that we shall be unfair to the theory if we confine our attention to the trans-subjective variety of it and if we assume that moral feeling is of the nature of emotion rather than sensation.

I shall begin with some general remarks about emotion.

(i) An emotion, e.g., an experience of fearing or hating, as distinct from an emotional mood, such as a state of apprehension or of crossness, is always directed to a cognised object. This may be real or hallucinatory, e.g., one may be afraid of a real man who is pointing a revolver at one or of an hallucinatory appearance of such a man in a dream. Again, if the object be real, it may be correctly or more or less incorrectly cognised, e.g., one may be afraid of a real physical object which one sees when crossing a field in twilight and takes to be a man pointing a revolver at one, and this object may really be a harmless scarecrow.

(ii) We must distinguish between what I will call "mediated" and "unmediated" emotions. Sometimes when a person feels a certain emotion towards a certain object he has an experience which may be described as feeling that emotion towards that object *in respect of* certain characteristics which he believes (rightly or wrongly) that it possesses. In that case I shall say that his emotion is *mediated* by this belief about the characteristics of the object, and I shall call these characteristics the "mediating characteristics" of the emotion. Often, however, the emotion is not felt in respect of any characteristic which the experient believes the object to have. In that case I shall say that the emotion is *unmediated*. If I am angry with a person, e.g., I may feel this anger in respect of some fault which I believe (rightly or wrongly) that he has committed. But I may feel angry with a person, and still more obviously I may dislike him, just directly and, as we say, "for no assignable reason." This is an example of an ø emotion.

(iii) Presumably every occurrence of any emotion, whether mediated or unmediated, has a total *cause*. In many cases, no doubt, an essential factor in that cause is the presence of certain characteristics in the object. I will call these "evoking characteristics." In the case of a mediated emotion the evoking and the mediating characteristics may be, and no doubt often are, wholly or partly the same. But very often they must be different; for the object often does not really have the characteristics which the experient believes it to have and in respect of which he feels his emotion towards it.

(iv) It is commonly held that certain kinds of emo-

tion are in some sense "appropriate to" objects which have certain characteristics, and that they are "inappropriate to" objects which lack these or which have certain others, e.g., fear is held to be appropriate only to objects which are dangerous. Again, it is held that for a given degree of dangerousness there is, within fairly narrow limits, a fitting degree of fear. To fear objects which are not really dangerous is described as "irrational"; and to fear intensely objects which are only slightly dangerous is described as "inordinate."

It is a well-known fact that if a person begins by feeling an unmediated emotion towards an object he is very liable to go on to ascribe to that object such characteristics as would make the emotion appropriate and to ascribe to those characteristic ssuch a degree as would make his emotion ordinate. A very familiar example of this is provided by persons who are jealous of others. Lastly, if a person feels a mediated emotion towards an object in respect of a characteristic to which that emotion is inappropriate, he is very liable to divert his attention from this fact and to ascribe to the object another characteristic in respect of which the emotion would be appropriate. These tendencies, which have been perfectly familiar to playwrights, preachers and plain men throughout the ages, have been hailed as great discoveries of modern psychology under the name of "rationalisation."

We are now in a position to consider the trans-subjective dispositional form of the Moral Sense Theory. In essence the theory is that such judgments as "That act is right (or is wrong)" are analogous to such judgments as "That food is nice (or is nasty)." The correct analysis of them is some variant on the formula "That act would evoke a moral pro-emotion (or anti-emotion) in any human being who might at any time contemplate it." There might have to be qualifications about the individual being "normal" and being "in a normal state," but we need not trouble about them at present.

Now this form of the theory does avoid the first objection which I made against the naively realistic form of it. It does not have to assume that one person literally has knowledge by acquaintance of the intentions of another. It does not have to assume that the experience of having a moral feeling when contemplating an act of one's own is fundamentally different in kind from that of having a moral feeling when contemplating a similar act of another person. For we can and do have emotions towards objects which are cognised only conceptually, and we can and do feel such emotions in respect of characteristics whose presence is only conceived and not perceived.

It seems to me that the main difficulties of the theory can be summed up in the following three questions: (i) Can it deal with the fact that judgments like "That act is right" seem always to be grounded upon the supposed presence in the act of some non-ethical right-inclining characteristic, such as being the fulfilment of a promise? (ii) If so, can it deal with the further fact that the connexion between a right-inclining characteristic and the rightness which it tends to convey seems to be necessary and synthetic? And (iii) can it deal with the fact that it seems not only intelligible but also true to say that moral pro-emotion is felt towards an act in respect of the characteristic of *rightness* and moral anti-emotion in respect of the characteristic of *wrongness*? I shall take these three questions in turn.

(i) I think that a fairly plausible answer, so far as it goes, can be made to the first question. We shall have to say that the right-inclining characteristic which is the ground of the judgment "That act is right" just is the mediating characteristic of the moral pro-emotion which is felt towards such acts. To say that every moral judgment is founded upon some non-ethical characteristic of the act which is its subject will be equivalent to saying that every moral emotion is a mediated emotion. Such characteristics as being an act of promise-keeping will be mediating characteristics for moral pro-emotion; such characteristics as being an act of lying or of deliberate cruelty will be mediating characteristics of moral anti-emotion.

It should be noticed that the theory can account quite plausibly for the facts which Ross describes under the head of his distinction between "*prima facie* duties" and "a duty proper." (I prefer to use the phrases "components of obligation" and "resultant obligation.") An act is known or believed to have various characteristics, e.g., to be an act of truth-telling, a breach of confidence, and an optimific act. The first and the third of these features give rise to components of obligation of various degrees of urgency towards doing it; the second gives rise to a component of a certain degree of urgency against doing it. According to circumstances the resultant obligation may be to do it or to avoid doing it. Now it is a perfectly familiar fact that an object may have several characteristics, and that it may call forth an emotion of one kind in respect of some of them and an emotion of the opposite kind in respect of others; so that the emotion towards the object as a whole may be predominantly of the opposite kind. The present theory would say that we tend to feel a moral pro-emotion of a certain strength towards the act in respect of its being one of truth-telling and in respect of its being optimific; that we tend to feel a moral anti-emotion of a certain strength towards

it as being a breach of confidence; and that our moral emotion towards it as a whole is the resultant of these two tendencies, and may be either predominantly *pro* or predominantly *anti* according to circumstances.

(ii) The second question is much harder. It is alleged, e.g., that the proposition "Any act of promise-keeping tends as such to be right, and any act of promise-breaking tends as such to be wrong" is necessary, self-evident, and synthetic. On the present theory of deontic judgments this would be equivalent to something like the following proposition: "It is necessary, self-evident and synthetic that any human being who should contemplate an act which he believed to be one of promise-keeping would tend to feel a moral pro-emotion towards it, and that he would tend to feel a moral anti-emotion towards any act which he believed to be one of promise-breaking."

Now it might be objected that the latter statement is certainly false. It is a purely contingent fact that human beings have a disposition to feel moral emotions at all. They might have been as devoid of them as they are of a disposition to have special sensations in presence of magnets. Moreover, granted that they do have such an emotional disposition, it is a purely contingent fact that moral emotions are mediated in the particular ways in which they are. It is quite conceivable that the belief that an act is one of promise-keeping should have mediated a moral *anti*-emotion, and that the belief that it is one of promise-breaking should have mediated a moral *pro*-emotion; just as it is conceivable that men should have liked the taste of castor oil and disliked that of sugar. In that case, on the present theory, promise-breaking would have tended to be right and promise-keeping to be wrong; just as castor oil would have been nice and sugar nasty.

So the objection comes to this. If the present form of the Moral Sense Theory were true, certain propositions which are in fact necessary and knowable *a priori* would have been contingent and knowable only empirically. Therefore the theory is false. I am sure that this is the most important of Price's objections to the Moral Sense Theory, though I have developed it in my own way. What are we to say about it?

It is plain that there are only two lines of defence open to the present form of the Moral Sense Theory. (a) One is to argue that propositions like "Any act of promise-keeping tends as such to be right" are *not* necessary. (b) The other is to argue that propositions like "Any human being who should contemplate an act which he believed to be one of promise-keeping would tend to feel a moral pro-emotion towards it" are *not* contingent. Let

us consider the two alternatives in turn:

(a) I think that this line of argument would divide into two parts, which might be called the "offensive" and the "defensive." The offensive part would take the opposite view as a hypothesis and try to show that it is untenable. The defensive part would try to explain why certain propositions which are in fact empirical and contingent appear to many people to be *a priori* and necessary.

(a, α) The offensive part may be put as follows: What precisely do our opponents maintain? If we may take Price as their ablest representative, they seem to assert something like the following doctrine. Suppose that a person reflects, e.g., on the situation of being asked a question and on the notions of responding to it by a true answer and responding to it by a false answer. Then he will find it self-evident that the former kind of response has a certain relation of "moral fittingness" and that the latter has an opposite relation of moral unfittingness" to such a situation. This relation of moral fittingness or unfittingness is held to be unique and unanalysable. And the process of recognising that it necessarily holds between certain kinds of response and certain kinds of situation is held to be analogous to that of recognising that certain mathematical terms, e.g., stand in certain mathematical relations.

Now the objection which will be made by supporters of the Moral Sense Theory is twofold. It will be said that the doctrine just enunciated involves *a priori* concepts and synthetic *a priori* judgments, and that neither of these is admissible. We will take these two points in order.

If there is a simple unanalysable relation of moral fittingness or unfittingness, it is certainly not manifested to us by any of our senses. We literally *see* that one coloured patch is surrounded by another; we literally *hear* that two notes, sounded together or in very close succession, concord or discord with each other; and so on. In such cases we presumably derive our ideas of the relation of *surrounding* and the relation of *concording or discording* by comparison and abstraction from such sensibly presented instances of terms standing in these relations. It is plain that we do not acquire the idea of moral fittingness or unfittingness in this way. Nor do we derive the idea from instances of terms presented to us by introspection as standing in that relationship. Introspection presents us with certain of our own experiences as standing in certain temporal relations, e.g., as being in the same specious present and partly overlapping in time, and so on. Again, since the relation of moral fittingness or unfittingness is held to be simple and unanalysable,

the idea of it cannot be one which we have constructed in thought from elements presented separately or in different contexts by sensation or introspection or both. (The idea of the complex relationship of a colonel to the subordinate officers of his regiment, e.g., is no doubt reached in some such way as this.) But it is held by many philosophers to be a fundamental epistemological principle that every idea is either derived by abstraction from instances presented in sensation or introspection or is an intellectual construction from elements so derived. If this principle be admitted, it is impossible that we should have any conception of the relations of moral fittingness and unfittingness as described by such moralists as Price.

For my part I attach very little weight to this argument. I can see nothing self-evident in what I will call for short "Hume's Epistemological Principle," and I am not aware that any conclusive empirical evidence has been adduced for it. It seems to me to be simply a useful goad to disturb our dogmatic slumbers, and a useful guide to follow until it begins to tempt us to ignore some facts and to distort others. I am inclined to think that the concepts of Cause and of Substance are *a priori* or contain *a priori* elements; at any rate I have never seen any satisfactory account of them in accordance with Hume's Principle.

The second point in the offensive part of the argument is this: Suppose, if possible, that "right" and "wrong" are simple unanalysable notions, as Price, e.g., held them to be. Then any proposition which asserts a connexion between some non-ethical characteristic, such as promise-keeping, and tendency to be right must be synthetic. Now a proposition may be synthetic and contingent or analytic and necessary, but it is an admitted general principle that no proposition can be both synthetic and necessary. Therefore the combined doctrine that "right" and "wrong" are unique unanalysable notions and that "such propositions as" Any act of promise-keeping tends as such to be right" are necessary must be false.

Such an argument would have different effects on different persons. Suppose that A and B are both quite convinced up to a certain moment of the truth of a certain general principle, and suppose that at that moment C brings to their notice an apparent counter-instance. If each is to be self-consistent, *something* will have to give way in each of them. But it need not be the same something. A may remain completely certain of the general principle; he will then have to maintain that the instance is only apparently contrary to it and explain why it seems to be so. B may find it impossible to doubt that the instance is contrary; he will then be forced to give up the general principle and explain why it seemed evident.

These are the two extreme possibilities. Between them are numberless possible intermediate alternatives, where the person concerned is led to feel *some* doubt of the unqualified truth of the principle and *some* doubt whether the apparent counter-instance really conflicts with it. Speaking for myself, I occupy one of these intermediate positions. As for Price, he would have been completely unmoved by this kind of argument. For he held, in full knowledge of Hume's doctrine and in conscious opposition to it, that there are plenty of synthetic necessary facts in other departments beside that of morals. For these reasons I think that it is rather futile to rely on a general argument of this kind.

(a, β) The defensive part of the argument might take the following line. Civilised men throughout human history have been assiduously conditioned in infancy and youth by parents, nurses, schoolmasters, etc., to feel moral pro-emotions towards acts of certain kinds and to feel moral anti-emotions towards acts of certain other kinds. Moreover, if we consider what kinds of acts are the objects of moral pro-emotions and what kinds are the objects of moral anti-emotions we notice the following facts about them. The former are acts whose performance by most people on most occasions when they are relevant is essential to the stability and efficient working of any society. The latter are acts which, if done on many occasions and by many people, would be utterly destructive to any society. On the other hand, the former are acts which an individual is often strongly tempted to omit, and the latter are acts which he is often strongly tempted to commit. This is either because we have strong natural impulses moving us to omit the former and to commit the latter, or because the attractive consequences of the former and the repellent consequences of the latter are often remote, collateral, and secondary. It follows that any group of men in which, from no matter what cause, a strong pro-emotion had become associated with acts of the first kind and a strong anti-emotion with acts of the second kind would be likely to win in the struggle for existence with other groups in which no such emotions existed or in which they were differently directed. Therefore it is likely that most of the members of all societies which now exist would be descendants of persons in whom strong moral pro-emotions had become attached to acts of the first kind and strong anti-emotions to acts of the second kind. And most existing societies will be historically and culturally continuous with societies in which such emotions-had become attached to such acts. These causes, it might be argued, conspire to produce so strong an association between such emotions and such acts in most members of every existing society that the connexion between the emotion and the act seems to each individual to be necessary.

No doubt this line of argument will produce different effects on different persons. For my own part I am inclined to attach a good deal of weight to it.

(b) I pass now to the second kind of defence which might be made for the dispositional form of the Moral Sense Theory. This is to contend that the proposition about human emotional dispositions which, according to the theory, is equivalent to "Any act of promise-keeping tends as such to be right" *is* necessary. It might be thought that this contention is so palpably absurd as not to be worth putting forward. But I believe that a case can be made for it, and I propose to make it.

We must begin by noting that the proposition which is equivalent to "Any act of promise-keeping tends as such to be right" could not with any plausibility be taken to be the crude unqualified proposition "Any human being has a disposition to feel a moral pro-emotion whenever he contemplates an act which he believes to be one of promise-keeping." So far from being necessary the latter proposition is not even true. To make it true it will have to be qualified somewhat as follows. We must substitute for it the proposition "Any *normal* human being has a disposition to feel a moral pro-emotion towards any act which he believes to be one of promise-keeping if he contemplates it when he is in a normal state."

Now it might be argued that, when the proposition is thus qualified, it *is* necessary. For, it might be said, it has then become *analytic*. It is part of the definition of a "normal" human being that he has a disposition to feel moral emotion, and that he will feel that emotion in its pro- form towards acts which he believes to be ones of promise-keeping, of truth-telling, of beneficence, and so on. And it is part of the definition of "being in a normal state" that when one is in such a state this moral-emotional disposition will not be inhibited altogether or excited in abnormal ways.

No doubt the immediate answer which an opponent of the Moral Sense Theory would make to this contention is the following: He would say that such propositions as "Any act of promise-keeping tends as such to be right" are not only necessary but *synthetic*. The defender of the dispositional form of the Moral Sense Theory has shown that, on his analysis, they would be necessary only at the cost of showing that they would be *analytic*. This answer is correct so far as it goes, but I think that the defender of the Moral Sense Theory could rebut it as follows.

The fact is that it is often by no means easy to say whether a proposition is analytic or not. The analytic propositions of real life are not like the trivial examples

in logic-books, such as "All negroes are black" or "All right angles are angles." The following are much better worth considering, e.g., "The sun rises in the east," "A freely suspended magnet sets itself with its axis pointing north and south," and "Pure water boils at 100 °C under a pressure of 76 centimetres of mercury." The first of these is analytic if "east" and "west" are defined by means of the sun, and synthetic if they are defined by means of the magnetic or the gyroscopic compass. The second is analytic if "north" and "south" are defined by means of the magnetic compass, and synthetic if they are defined by means of the sun or the gyroscopic compass. The third might be taken as a definition of '100 °C. But if that term were defined in some other way, e.g., thermodynamically, as on Lord Kelvin's absolute scale, it might be regarded as an analytic proposition about *pure water*. For an important element in the definition of "pure water" is that it has a certain boiling-point under certain standard conditions.

Two important points emerge from these examples. The first is that the same type-sentence may express both an analytic and a synthetic proposition, and that a person who uses several tokens of this type even in a single discourse may sometimes be expressing the analytical and sometimes the synthetic proposition. The former is necessary and the latter is contingent. It would not be surprising if a person should sometimes become confused in such cases and think that every token of this type expresses *one and the same proposition* which is *both* synthetic and necessary.

The second point is this. Such an analytic proposition as "Pure water boils at 100 °C under a pressure of 76 centimetres of mercury" has at the back of it a whole system of interconnected empirical generalisations, apart from which it would never have been worth anyone's while to formulate it. It would take me far too long even to begin to state a few of these empirical generalisations. It will suffice to say that they are all represented in the various qualifications which make the proposition "Pure water boils at 100 °C. under a pressure of 76 centimetres of mercury analytic.

Now it might be suggested that facts like these throw some light on the alleged synthetic necessity of such propositions as "Any act of promise-keeping tends as such to be right," and on the claim of defenders of the dispositional form of the Moral Sense Theory that the equivalent propositions about human emotional dispositions are necessary because analytic.

The proposition "Any act of promise-keeping would tend to call forth a moral pro-emotion in any *normal* human being who might contemplate it when in a *nor-*

mal state" is obviously rather like the proposition "Any sample of pure water boils at 100 ºC under the *normal* atmospheric pressure, i.e., 76 centimetres of mercury." Just as the latter is analytic, but is founded on a whole mass of interconnected empirical generalisations, so is the former. I will now try to justify this statement.

It is an empirical fact that the vast majority of men have a disposition to feel moral emotions, and that the minority who lack it differ in many other ways from the majority of their fellows. It is an empirical fact that there is very substantial agreement among men in the kinds of act which call forth moral pro-emotion and in the kinds which call forth moral anti-emotion. The small minority of men who habitually feel moral pro-emotion where most of their fellows feel moral anti-emotion, or vice versa, are generally found to be odd and abnormal in many other ways. There is, in fact, so high a degree of positive association between moral and non-moral normality that it would make very little difference in practice whether we defined a "normal" man solely by reference to his moral dispositions or solely by reference to his non-moral dispositions, or by reference to a mixture of both. But the proposition that any normal human being would tend to feel a moral pro-emotion towards any act which he believed to be one of promise-keeping would be synthetic if one defined "normality" solely by reference to non-moral dispositions, whilst it might well be analytic if one defined it wholly or partly in terms of moral dispositions.

Again, there is a very high degree of positive association between the tendencies to feel moral pro-emotion towards acts of promise-keeping, of truth-telling, of beneficence, etc.; and there is perhaps an even stronger degree of positive association between the tendencies to feel moral anti-emotion towards acts of treachery, of unfairness, of cruelty, etc. Therefore it would make little practical difference which of these mediating characteristics was included and which was omitted from the definition of "normality." Now, if the tendency to feel moral pro-emotion towards any act which is believed to be one of promise-keeping were included in the definition of "normality," the proposition that any normal man would tend to feel such an emotion towards such acts would be analytic; whilst, if this were omitted and "normality" were defined by reference to some of the other mediating characteristics of moral emotion, this proposition would be synthetic.

It therefore seems likely that, if the analysis which the dispositional form of the Moral Sense Theory offers for such propositions as "Any act of promise-keeping tends as such to be right" were correct, a sentence of this

type might often express a proposition which is analytic and necessary and might as often express one that is synthetic and contingent. If so, it is not unlikely that a confusion should arise and that it should be thought that *every* such sentence expresses one and the same proposition which is *both* necessary and synthetic.

It remains to say something of the qualification "when in a normal state," which has to be added to make the statement universally true, and which at the same time makes it more nearly analytic. It may be compared to the qualifications about the water being pure and the barometric pressure being normal in my example about boiling-point.

At the back of this qualification lie certain negative and certain positive empirical facts. It is found that a person who generally does feel moral pro-emotions towards acts of certain kinds and moral anti-emotions towards acts of certain other kinds will on some occasions not do so. He may feel no moral emotion; or perhaps on very exceptional occasions the normal form of his moral emotion may be reversed. These are the negative facts. The positive facts are certain empirical generalisations about the kinds of occurrent conditions under which such inhibitions or reversals of moral emotion tend to take place. "Being in a normal state" is then defined in terms of the absence of such conditions, e.g., not being angry with or jealous of or frightened by the agent whose act is being contemplated. Now, although one has at the back of one's mind a fairly adequate but rather confused idea of these negative conditions, only one or two of them will be explicitly before one's mind on any particular occasion when one uses the expression "in a normal state." According as one or another is in the foreground on a given occasion the same sentence may express an analytic or a synthetic proposition.

I suggest, then, that defenders of the dispositional form of the Moral Sense Theory might attempt in some such ways as these to rebut the objection that, whilst propositions like "Any act of promise-keeping tends as such to be right" are *necessary and synthetic*, the propositions which it asserts to be their equivalents are either *contingent* or *analytic*.

(iii) The third difficulty which the Moral Sense Theory, in the form of it which we are considering, has to meet is this. It might be alleged that the mediating characteristics in respect of which a person feels moral pro-emotion or anti-emotion towards an act which he contemplates are the supposed *rightness or wrongness* of the act. Suppose, e.g., that a person feels a moral anti-emotion when he contemplates an act which he believes to be one of promise-breaking. Then, it might be said, he

does so only in so far as he believes promise-breaking to be wrong. Suppose that he believed the act to be one of promise-breaking but did not believe that such acts tend to be wrong. Then, it might be alleged, there is no reason to think that he would feel a moral anti-emotion towards it.

Let us begin by considering what view a Rationalist, like Price, would take on this question of the mediating characteristics of moral emotion. I think that the following is a fair statement of his position. It is a necessary proposition that any rational being who contemplated an act which he believed to be one of promise-breaking would tend to feel towards it a moral anti-emotion. But, though true and necessary, it is not self-evident. It is a logical consequence of two more fundamental propositions, each of which is self- evident. They are these: (a) It is self-evident to any rational being that any act of promise-breaking tends as such to be wrong. (b) It is self-evident that any rational being who contemplated an act which he believed to be wrong would feel towards it a moral anti-emotion.

We have already considered what the supporters of the Moral Sense Theory might say about the first of these propositions. What are we to say about the second? It seems to me that everything depends here on how much we put into the connotation of the phrase "rational being." On a narrower interpretation of that phrase proposition (b) is synthetic but contingent, on a certain wider interpretation that proposition becomes necessary but analytic. Sometimes the one interpretation and sometimes the other is at the back of one's mind without one realising the fluctuation, and so one is inclined to think that proposition (b) is both necessary and synthetic.

A "rational being," on the narrowest interpretation, means roughly one who is capable of comparing, abstracting, and forming general notions; who is capable of seeing necessary connexions and disconnexions between terms and between propositions; and who has the power of making inferences, both deductive and inductive. I call this the "narrowest" interpretation, because it takes account only of cognitive characteristics and leaves out emotional and conative ones. The next stage in widening it would be to include in the definition of a "rational being" what I will call "purely intellectual" emotions and conations, e.g., intellectual curiosity, taking pleasure in neat arguments and displeasure in clumsy ones, desire for consistency in one's beliefs, and desire to apportion the strength of one's beliefs to the weight of the evidence.

Let us say that a person who had the cognitive, conative and emotional dispositions which I have just enumerated would be rational "in the ethically neutral

sense." Suppose that Price were correct in thinking that moral fittingness and unfittingness are relations which hold of necessity between certain types of response and certain types of situation. Then a person who was rational in the ethically neutral sense would in principle be capable of having ideas of right and wrong and of making moral judgments. (I say "in principle" because (a) he would, by definition, have the *general* capacity to see necessary connexions between terms and between propositions, whilst (b) it might happen that his insight in this particular department was lacking, as that of some rational beings is in the department of mathematical relations.) But, so far as I can see, there would not be the slightest inconsistency in supposing that a being who was rational in the ethically neutral sense, and did in fact have the ideas of right and wrong and make moral judgments, was completely devoid of specifically moral emotion and conation. The fact that he knew or believed A to be right and B to be wrong might arouse in him neither moral pro-emotion towards the former nor moral anti-emotion towards the latter, and it might not evoke in him the slightest desire to do A or to avoid doing B or vice versa. I cannot see any *logical* impossibility in the existence of such a being; whether it would involve a conflict with some of the de facto laws of psychology I do not know.

Now the vast majority of the beings whom we know to be rational in the ethically neutral sense do in fact feel moral pro-emotion towards acts which they believe to be right and moral anti-emotion towards those which they believe to be wrong, and they are in fact to some extent attracted towards doing the former and repelled from doing the latter. Moreover, it is *logically* impossible that these specifically moral emotions and desires should exist in a being who was not rational in the ethically neutral sense; for their characteristic objects can be presented only by a process of reflective thinking. The wider interpretation of the phrase "rational being" includes these specifically moral conative and emotional characteristics in addition to those which constitute the definition of "rational" in the ethically neutral sense. It is, of course, logically impossible that a person who is rational in this widest sense should fail to feel moral pro-emotion towards what he believes to be right and moral anti-emotion towards what he believes to be wrong. But this is a merely analytical proposition. It is synthetic and contingent that a person who is rational in the ethically neutral sense should be so in the wider ethical sense also. But the fact that rationality in the ethically neutral sense is almost invariably accompanied in our experience by the additional features which convert it into ethical rationality and the fact that the latter logically entail the former produce a confusion in our minds. We are thus

led to think that the proposition that any rational being would feel a moral pro-emotion towards any act which he believed to be right and a moral anti-emotion towards any that he believed to be wrong is both necessary and synthetic.

So much for the Rationalist account of moral emotion and its mediation by the characteristics of rightness and wrongness. What can the Moral Sense Theory, in its trans-subjective dispositional form, make of the alleged facts?

On the face of it this theory is presented with the following difficulty. Suppose that we try to combine the alleged fact that rightness and wrongness are the mediating characteristics for moral emotion with the analysis of moral judgments given by the theory in question. Then we seem to be committed to the following proposition: "A person will tend to feel a moral anti-emotion towards an act which he believes to be one of promise-breaking so far and only so far as he believes that most persons when in a normal condition would feel such an emotion in contemplating such an act." Now this has a *prima facie* appearance of circularity; and, even if it be neither logically nor causally circular, it certainly does not seem very plausible.

The first remark that I have to make is that the objection just stated rests on a premise which is plausible but false. It tacitly assumes that, if the correct analysis of the proposition "S is P" is "S is p_1-and-p_2," then anyone who is believing the former proposition is *ipso facto* believing the latter. Now there may be some sense of "believe" in which this is true; but there certainly is an important sense in which it is false. It is quite obvious that a number of persons who accept different and incompatible analyses of a proposition may all believe it; and therefore there must be a sense in which some at least of them believe it without *ipso facto* believing the proposition which is its correct analysis. This is particularly obvious in the present case. Nearly everyone believes that acts of promise-breaking tend as such to be wrong; but some of these persons think that wrongness is a simple characteristic, others think that it can be analysed in one way, and others think that it can be analysed in various other ways. So, even if the correct analysis of "X is wrong" is "Any normal person who should contemplate such an act as X when in a normal state would feel a moral anti-emotion towards it," it does not follow that the correct analysis of "A believes that X is wrong" is "A believes that any normal person who should contemplate such an act as X when in a normal state would feel a moral anti-emotion towards it." So it is not fair to say that the Moral Sense Theory must hold that anyone

who feels a moral anti-emotion towards an act in respect of his belief that it is wrong is *ipso facto* feeling that emotion in respect of his belief that any normal person would feel such an emotion if he were to contemplate such an act while in a normal state.

I suppose that this argument would be generally admitted as applied to the case of a person who did not accept, or did positively reject, the analysis of moral judgments proposed by the Moral Sense Theory. But it might be said that it will not apply to the case of a person who accepts that analysis. I think, however, that even this could be questioned. A person may have assented to a certain analysis of a proposition when the question of its analysis and the arguments *pro* and *con* were before his mind. He may continue to accept it, in the *dispositional* sense that he *would* assent to it again at any time when the question was raised for him. But during the intervals he may often have the experience of believing the proposition without thinking of the analysis of it which he has accepted. Therefore it seems to me that even an adherent of the Moral Sense Theory might often feel a moral anti-emotion towards an act in respect of his belief that it is wrong without *ipso facto* feeling that emotion in respect of the belief that it has those characteristics which he holds to be the correct analysis of "being wrong."

So much for the dialectics of the matter. But what is really happening when a person is said to feel a moral pro-emotion or anti-emotion towards an act in respect of his belief that it is right or that it is wrong? We must begin by distinguishing what I will call "first-hand" and "second-hand" emotion. Suppose that a certain word has been very often used in connexion with objects towards which a certain kind of emotion has been felt and that it has seldom or never been used except on such occasions. Then this word may come to act as a stimulus calling forth this kind of emotion. When the emotion is evoked in this way I call it "second-hand."

Now there is no doubt that a great deal of moral emotion is, in this sense, second-hand. And there is no doubt that the words which have come by association to act as evokers of second-hand moral emotion are the words "right" and "wrong." When a person is said to feel a moral emotion towards an act in respect of his belief that it is right or that it is wrong what is really happening is very often the following. He knows or believes that acts of this kind are commonly *called* "right" or called "wrong." He repeats these words *sotto voce* to himself or has auditory images of them when he thinks of the act in question; and by association they evoke a second-hand moral pro-emotion or anti-emotion towards the act. Plainly there is nothing in this to cause

difficulty to the supporters of the Moral Sense Theory.

But of course this does not cover the whole field. There is first-hand moral emotion; indeed, if no one had ever felt a first-hand emotion of a given kind, it is difficult to believe that anyone could now feel a second-hand emotion of that kind. What is happening when a person is said to be feeling a *first-hand* moral emotion towards an act in respect of his belief that it is right or that it is wrong? I can give only a very tentative answer to this question, based on my own imperfect introspection of a kind of situation with which I am not very familiar.

It *seems* to me that in such cases I do not first recognise or think that I recognise a quality or relation of rightness or wrongness in the act, and *then* begin to feel a moral pro-emotion or anti-emotion towards it in respect of this knowledge or belief. What I seem to do is to consider the act and its probable consequences under various familiar headings. "Would it do more harm than good? Would it be deceitful? Should I be showing ingratitude to a benefactor if I were to do it? Should I be shifting onto another person's shoulders a burden or a responsibility which I do not care to bear for myself? "In respect of each of these aspects of the act and its consequences I have a tendency to feel towards the act a certain kind of moral emotion of a certain degree of intensity. These emotional dispositions were largely built up in me by my parents, schoolmasters, friends and colleagues; and I know that in the main they correspond with those of other persons of my own nation and class. It seems to me that I call the act "right" or "wrong" in accordance with my final moral-emotional reaction to it, after viewing it under all these various aspects, and after trying to allow for any permanent or temporary emotional peculiarities in myself which may make my emotional reaction eccentric or unbalanced. By the time that this has happened the features which I had distinguished and had viewed and reacted to separately have fallen into the background and are again fused. They are the real mediating characteristics of my moral pro-emotion or anti-emotion; but I now use the omnibus words "right" or "wrong " to cover them all, and say that I feel that emotion towards the act in respect of my belief that it is right or that it is wrong.

C.D. BROAD ON MORAL SENSE THEORIES IN ETHICS

Commentary by Robert Cowan

ROBERT COWAN is a
Lecturer in Philosophy at
the University of Glasgow.
His research is focused on
the intersection between
ethics, epistemology,
and the philosophy of
mind. He has recently
published papers on ethical
intuitionism and the nature
and epistemology of ethical
perception. He is currently
writing about perceptual
theories of the emotions
and conceptualist accounts
of substantive a priori
ethical knowledge, and is
also co-editing (with Anna
Bergqvist) a collection
of papers on Evaluative
Perception.

C.D. BROAD ON MORAL SENSE THEORIES IN ETHICS

Commentary by Robert Cowan

1. See also Frankena 1955 and Raphael 1947. For a twenty-first century defence see Filonowicz 2008.

C.D. BROAD'S 1945 *Reflections* (as I will call it) stands out as one of the few serious examinations of Moral Sense Theory in twentieth century analytic philosophy.[1] It also constitutes an excellent discussion of the interconnections that allegedly exist between questions concerning what Broad calls the 'logical analysis' of moral judgments and questions about their epistemology. In this paper I make three points concerning the interconnectedness of the analytical and epistemological elements of versions of Moral Sense Theory. First, I make a general point about Broad's association between the Naïve Realist Moral Sense Theory (an epistemological view) and Objectivist Moral Sense Theory (a 'logical analysis'). Second, I raise doubts about one of Broad's arguments that Trans-Subjectivist Moral Sense Theory (logical analysis) can account for the apparent synthetic necessity of general moral propositions (epistemological). Third, I briefly discuss a view about logical analysis that should be of interest to contemporary Moral Sense Theorists – Neo-Sentimentalism – and respond to an argument whose conclusion is that this analysis is incompatible with a particular kind of epistemological view.

I.

Broad says that all epistemological Moral Sense Theories endorse the view that there is a

> peculiar kind of experience which human beings are liable to have when they contemplate certain acts, and that this can take either of two opposite forms, viz., a pro-form and anti-form… this experience is of the nature of feeling, where 'feeling' is used to include both sensation and emotion as distinguished from thought."(145)

Such experiences are allegedly responsible both for the formation of fundamental moral concepts, e.g., *moral rightness*, and the formation of singular moral judgments, e.g., 'that is right'.

Some clarification is in order. By 'experience' I take Broad to be picking out a class of mental items that have *phenomenal character*, that we are to some extent *passive* in the face of, and which aren't *based* upon other mental items. Visual experience is a paradigm (inten-

tional) example. By alleging that moral experience is 'peculiar' Broad is following Moral Sense Theorists like Hutcheson (1969a and 1969b) who seemed to think that moral experiences – approbation and disapprobation – had a distinctive *character* and *aetiology*.

Regarding character: the approbation experienced when contemplating a benevolent action allegedly feels different from that had in response to a beneficial, self-interested action. Hutcheson took moral ideas (roughly: concepts) to be simple, i.e., not compounded out of simpler ideas. Given Hutcheson's Lockean epistemology, moral ideas must therefore be direct 'copies' from a distinctive kind of experience. Further, simple ideas must be of properties rather than relations. Regarding aetiology: moral experiences aren't the product of the canonical sensory modalities, e.g., vision. Hutcheson famously talks explicitly of a distinctive moral sense although his criterion for 'sense' is permissive: "a determination of the mind, to receive any idea from the presence of an object which occurs to us, independent of our will" (Raphael 1947: 264-5). This is compatible with the moral sense being a mere disposition to undergo experiences, given certain stimuli. However, at other points, Hutcheson directly compares the moral sense and the canonical senses which might make us suspect him of positing a dedicated moral faculty, or at any rate, a faculty which has a distinctive 'moral' component.

Although Broad may not have noticed this, the specification that the moral experience is 'of the nature of "feeling"' is required to distinguish Moral Sense Theory from a version of Rational Intuitionism which conceives of moral experiences as intellectual experiences ('intuitions') that are in some way produced by our faculty of Reason, i.e., phenomenal states that are non-voluntary, baseless, etc.[2] Indeed, Hutcheson's own criterion for a sense is ambiguous between an empiricist and a rationalist construal. In passing, it is worth noting that in recent years philosophers (e.g. Dancy (forthcoming) and Kauppinen (2014)) have defended the view that 'moral intuitions' might actually be the manifestation of emotional dispositions.

Broad discusses two epistemological accounts available to Moral Sense Theorists: Naïve Realism and Dispositionalism.

Naïve Realism is the view that the moral sense is a power of perceiving non-dispositional moral properties. When the appropriate relation obtains between the moral sense and non-dispositional properties, the nature of those non-dispositional properties is revealed to subjects in experience or 'sensation' (147-8). According to Broad, Naïve Realism requires that subjects are acquainted with

2. For a recent defence see Chudnoff 2013.

3. I am assuming that Broad endorses something like Russell's view of acquaintance presented in chapter 5 of his *Problems of Philosophy* (1912).

non-dispositional moral properties, where 'acquaintance' appears[3] to pick out something that is *non-conceptual,* which is in some way *direct* (148-9) and facilitates experiences that *resemble* the relevant perceptual objects (143-4).

Dispositionalists deny that the moral sense is a power of perceiving non-dispositional moral properties. Instead, the moral sense is merely "an organ of emotional reaction" (Frankena 1955: p. 366) to the non-moral properties of actions (Broad seems to be focused on particular actions and act-types in *Reflections*). There is no acquaintance with non-dispositional moral properties, although there may be something worth calling acquaintance (non-sensory) with the phenomenal properties of the experience.

Regarding logical analysis, Broad seems to think that Moral Sense Theorists have two options. On the one hand they can endorse some kind of Objectivism, according to which moral judgments are about non-dispositional moral properties of actions or act-types. If Broad were following Hutcheson, he may have been assuming that the Naïve Realist thinks of moral properties as simple 'empirical' or 'natural' properties analogous to Moore's view that goodness was a simple non-natural property. Note, however, that this claim is optional. It seems open to Objectivist Moral Sense Theorists to reject this Lockean framework and argue that some properties – of which moral properties are an example – are neither simple (like yellowness) nor complex (a whole decomposable without remainder into simple parts), but are instead consequential properties.

Alternatively, Moral Sense Theorists could adopt a Trans-Subjectivist Dispositional view according to which moral judgments are analysable in terms of the dispositions of normal human agents to undergo the 'peculiar' moral experience (emotional or sensational) in response to the contemplation of actions or act-types.[4]

4. A possibility that Broad wasn't aware of is that emotional responses play a crucial role in fixing the reference of moral terms but moral properties are themselves *non-dispositional*. For this sort of view see Slote 2009: chapter 4.

I now focus on Broad's association of Naïve Realism with Objectivism. Certainly, it seems that a Naïve Realist Moral Sense Theorist is committed to some form of Objectivism: after all, on their view we are supposed to be acquainted (which is a 'success' relation) with non-dispositional moral properties.

However, Broad has quite powerful objections to Naïve Realism (I limit myself here to discussion of Broad's argument against Naïve Realism regarding the acts of other people). If moral sensing requires acquaintance, and if acquaintance requires that we stand in a direct relation to moral properties, then it is hard to see how there could be moral sensing. This is because we

are never directly aware of the putative objects of moral sensing – intentional actions – since we are not directly aware of intentions. At best, we can conceive of intentional actions (147-8). But that introduces an indirectness that is apparently antithetical to acquaintance. To this we might add that, if something like resemblance between experience and property were a condition on acquaintance, and if perception requires acquaintance, then perception of moral properties would seem to be a non-starter.[5] Hence, if Objectivist Moral Sense Theory hinged on Naïve Realism, this would make Dispositionalism (in both its analytical and epistemological forms) the only game in town for Moral Sense Theorists.

However, Objectivist Moral Sense Theory doesn't require that the moral sense be understood along the lines of Naïve Realism. Broad's discussion doesn't address what now seems a more attractive epistemological theory for Objectivist Moral Sense Theorists to adopt. Instead of claiming that the moral sense is a power of perceiving, where this requires acquaintance, Objectivist Moral Sense Theorists may instead hold the view that the moral sense is a power of perceiving moral properties which requires that subjects stand in a causal relation to those properties, which produces experiences (mental items that have phenomenal character, are passive, baseless) that (i) represent or present their instantiation, and, (ii) have some further property which is epistemologically relevant, e.g., reliability or special phenomenal character. Call this sort of view 'Representational Objectivism'.

There are two general ways this view could be developed.

Firstly, Representational Objectivists could develop an alternative account of moral sensing along Broad-inspired lines. According to Broad's full theory of the canonical sensory modalities – the complete statement of which can be found in his (1925) *The Mind and Its Place In Nature* – acquaintance is only a component of perception.[6] An overall perception also involves an 'external reference' which is in some sense 'based' (non-inferentially) upon sensation, but which can depict objects and properties that do not 'show up' in sensation. An example might be *a piece of gold*. One might therefore wonder whether a kind of moral sense theory could be developed whereby non-dispositional properties get depicted at the level of external reference. For example, consider Gilbert Harman's oft-quoted example of observing hoodlums setting fire to a cat. (Harman 1977) One might think that, in response to this, normal subjects could have moral perceptions (visual) which represent the *wrongness* of what the hoodlums are doing.[7] Perhaps, though, this runs against both the letter and

5. Broad is, however, never clear about what directness or resemblance require. Perhaps there are weaker conceptions of both of these which would make moral sensing more plausible.

6. Note, however, that Broad endorses a Sense-Datum theory, not Naïve Realism.

7. See e.g. Siegel 2010 on 'high-level' perception, and my discussion of moral perception (Cowan 2015).

spirit of Moral Sense Theory, which has (historically) involved an appeal to some kind of feeling response. Further, the view would presumably require supplementation to deal with moral judgments that are not responses to perceived instances.

This brings us to the second option for Representational Objectivists, which is to adopt a view according to which the moral experience is emotional. Indeed, this is something like the view held by some contemporary perceptual theorists (e.g. Tappolet 2012) of the emotions – according to which fear (for example) is a perception of danger, guilt a perception of wrongness, and so on – who can be thought of as defending a sort of Representational Objectivist Moral Sense Theory (although a Representational epistemology need not be paired with an Objectivist analysis[8]). To illustrate, consider again Harman's cat case. Representational Objectivists might argue that in this sort of case subjects have a perception whose external reference – perhaps this is a sort of seeming state, i.e., an experience with propositional content that has phenomenal character, is non-voluntary, baseless etc – is that the *hoodlums are setting fire to the cat*,[9] and have a moral experience of disapprobation in response to this which represents the *wrongness* of what they are doing. Perhaps the moral experience and external reference are in some way 'integrated' or mixed up with one another.[10]

Note that this view could accommodate emotional responses in response to cases that are not 'directly' perceived. One worry one might have about this sort of view is that it won't ground a comprehensive Moral Sense Theory of moral concept formation since the relevant emotional representation will be conceptual in character. However, philosophers like Tappolet think that emotions – including moral emotions – are non-conceptual, in which case they could play a crucial role in moral concept formation.

These comments are of course very general. I am not endorsing Representational Objectivist Moral Sense Theory. Instead, I am simply highlighting that Broad's objections to Naïve Realism shouldn't be taken as a reason to reject Objectivism (although there may be other good reasons for doing that).

II.

Broad goes on to defend a version of Moral Sense Theory which combines the Trans-Subjective Dispositional logical analysis and the Dispositional epistemological view. The most important objection that he grapples with is arguably the second (153-62). I'll briefly explain the objection and one of Broad's responses, raising some

8. Jesse Prinz (2004) defends something like a Representational epistemology allied with a Subjectivist analysis.

9. Broad apparently thought that our concept of *physical object* is an 'innate principle of interpretation'. This is the explanation for why physical objects show up in the external reference of perceptions. Perhaps one could make a similar point about *intentional actions*.

10. For this idea see Audi 2013.

11. I speak of 'pro tanto' reasons rather than employing Broad's misleading talk of 'tendencies' of acts to be right.

12. There may be some deadweight in Broad's characterisation of the Trans-Subjective Dispositional analysis: on plausible analyses of *dispositions*, they are manifested in *normal circumstances*. If that's right then there would be no need to make further reference to *normal circumstances* in the Trans-Subjectivist Dispositional analysis.

doubts about the persuasiveness of the latter.

The Trans-Subjective Dispositional analysis of the general moral proposition *promise-keeping is pro tanto right*[11] (hereafter 'moral analysandum') will amount to something like *normal people are disposed to feel a moral pro-emotion (approbation) when they contemplate acts of promise-keeping, in normal circumstances*.[12] However, whereas the moral analysandum is allegedly synthetic and necessary, the proposed Trans-Subjectivist analysans is contingent, and may indeed be false.

Broad's most intriguing response to this problem is to consider the claim that the Tran-Subjective analysans might be analytic and necessary. This would be true if the definition of *normal human being* made reference to dispositions to feel moral pro-emotions with regard to the relevant act-type(s), and *normal circumstances* were understood to be circumstances in which the disposition manifested.

However, even if this were plausible, Broad concedes that this won't satisfy those (Rationalists) who think that the moral analysandum is synthetic and necessary. In response Broad develops a complex argument (157-162), the broad thrust (pun intended) of which appears to be that (i) people are in error when they think that the moral analysandum is synthetic and necessary, and, (ii) this is precisely what we should expect if the Trans-Subjectivist Dispositionalist analysis were true. In more detail, here is my attempted reconstruction of the argument (the reader is of course invited to see where I might have gone wrong – at some crucial stages Broad is uncharacteristically opaque, see especially the last paragraph on p. 160):

> P1: If the Trans-Subjectivist analysis is true, then it can be expressed in tokens of the sentence type 'any normal human being has a disposition to feel a moral pro-emotion towards any act which he believes to be one of promise-keeping if he contemplates it when he is in a normal state'.

> P2: Tokens of the sentence type that express the Trans-Subjectivist analysans are ambiguous between a proposition that is analytic and necessary and one that is synthetic and contingent.

> P3: If P2 is true then tokens of the sentence type that express the moral analysandum will also be ambiguous between a proposition that is analytic and necessary and one that is synthetic and contingent.

> P4: If the moral analysandum is ambiguous between a proposition that is analytic and necessary and one that is synthetic and contingent then it is likely that people should become confused into

thinking that every instance of the sentence that expresses the moral analysandum expresses a proposition that is both synthetic and necessary.

C: If the Trans Subjectivist analysis of the moral analysandum is true then it is likely that people should become confused into thinking that every instance of the sentence that expresses the moral analysandum expresses a proposition that is both synthetic and necessary.[13]

Some clarification is in order.

P1 doesn't require explanation. Broad's reason for P2 is that, whether the relevant sentence token is analytic or synthetic depends upon whether we define (either partly or wholly) "Normal" relative to dispositions to have moral emotions or whether we do so purely in terms of non-moral dispositions. If the former, then the proposition will be analytic and necessary, if the latter, the proposition will be synthetic and contingent.

Broad's reasons for thinking that P3 is true are less clear. One way of interpreting him would be that he thought that in considering a token sentence expressing the moral analysandum we have before our minds one of the 'Trans-Subjectivist' propositions, i.e., either the analytic and necessary proposition (their proposed analysans) or the synthetic and contingent proposition. But that seems pretty implausible.

Another way of interpreting Broad would be that he thought that, if sentences expressing the Trans-Subjectivist moral analysans are ambiguous, the relevant sentence expressing the moral analysandum must also be. However, it is not clear that this is a reasonable inference to make. This is because the Trans-Subjectivist sentence is but one way of expressing the Trans-Subjectivist analysis – there are surely other sentences expressing the analysis that are not ambiguous. Surely Trans-Subjectivists think that those disambiguated sentences mean the same thing as the promise-keeping sentence, in which case it is not altogether clear what exactly the reason is for thinking that the promise-keeping sentence is itself ambiguous.

An alternative interpretation is that Broad thought that the points he made regarding the general liability for sentences to be ambiguous between analytic/necessary and synthetic/contingent propositions can find a direct application in the promise-keeping sentence. However, Broad doesn't provide an account of what propositions the promise-keeping sentence is supposed to be ambiguous between (instead he spends lots of time discussing how the Trans-Subjectivist sentence could be ambiguous). This seems to be an important lacuna in Broad's argument for P3.

13. An alternative interpretation is that Broad grants that the promise-keeping sentence expresses a synthetic and necessary proposition and is arguing that, although the Trans-Subjective analysans is analytic and necessary, sentences expressing it are liable to be confused for synthetic and necessary propositions. I don't interpret Broad this way as it seems to miss the point.

Here is a suggestion inspired by some of the things Broad says (see 157-8 and 160-1). Prior to considering the promise-keeping sentence, it is plausible that subjects will have an incomplete idea of the sorts of act-types that ground pro-tanto reasons. Perhaps John would only think about acts of injury, while Lucy thinks also of acts of gratitude etc. Neither thinks about promise-keeping. Given this, perhaps when such people consider the promise-keeping sentence, they think that it is non-definitional or synthetic. However upon considering different cases of promise-keeping – and having moral pro-emotions in response (let's suppose that they are normal and are in normal conditions) – and failing to find counterexamples, they might come to make the hasty judgment that all acts of promise keeping are pro-tanto right. Subjects may make the even hastier judgment that they *must* all be right. In doing so they might come to think that the proposition expressed by the sentence is synthetic and necessary, despite its either being synthetic and contingent (because promise-keeping is not included in the 'definition' of pro tanto moral reason) or analytic and necessary (because promise-keeping is included in the 'definition' of pro tanto moral reason).

I doubt that this helps matters. First, one might think that the suggestion is just plain implausible. Second, even if it were plausible, it would only have a limited purchase as it (i) won't apply to all subjects, and (ii) wouldn't explain why subjects to whom it does apply persist in thinking that the promise-keeping sentence expresses a proposition that is both synthetic and necessary. Hence, in my view, P3 is inadequately supported by Broad.

Regarding P4, it is worth considering Broad's nice account of why we might be confused regarding sentences expressing the Trans-Subjectivist analysans. Roughly he claims that the propositions expressed by tokens of the sentence-type 'Any normal human being...' are founded on a whole mass of interconnected empirical generalisations about moral and non-moral dispositions. Broad thinks these suggest a strong (though contingent) positive association between moral normality and non-moral normality. Given this, when people consider a token of the sentence type, the ideas of moral and non-moral normality may come and go from the mental "foreground" (161-2), creating the confusion that the proposition expressed is synthetic and necessary.

However, as was suggested in my discussion of P3, Broad doesn't provide a clear account of the propositions we are supposed to be confused between when considering sentences expressing the moral analysandum. Furthermore, and this is a general point: premises

of arguments which require that reflective interlocutors are in a persistent state of confusion about the status of a sentence invite the response: 'well I'm not bloody confused!' Hence, Broad's strategy here, although ingenious, doesn't strike me as likely to have much traction against opponents.

Although I'm unconvinced by this particular argument, this doesn't mean that the Trans-Subjectivist can't deal with the original objection, i.e., that their analysis is, at 'best', analytic and necessary, while the analysandum is synthetic and necessary. That said, the alternative strategies that Broad suggests (detailed in 154-58) don't seem likely to impress Rationalists either.

As a coda to this subsection it is worth very briefly mentioning that, although Broad provides something of a partial defence of Trans-Subjectivism in Reflections, he provides quite a scathing critique of it in his more famous work *Five Types of Ethical Theory* (1934). There he attributes a view like Trans-Subjectivism to Hume and claims that it has the consequence that moral disputes and moral inquiry are ultimately a matter of "experiment, observation, collection of statistics, and empirical generalisation". Broad thinks this is "simply incredible" (1934: 115).

One might think that Broad's developments of Trans-Subjectivism in the latter stages of Reflections might serve to deal with this sort of problem. However, even if the Trans-Subjective analysans is analytic and necessary, things are complicated by Broad's claim that the proposition is "founded on a whole mass of interconnected empirical generalisations" (159-60). Broad is unclear about what he means by "founded on", e.g., whether he thinks that this implicates an epistemic dependency of the analytic proposition upon the empirical generalisations, or something weaker like semantic dependency (to my mind it would involve both). He is, however, clear about some of the generalisations he has in mind, e.g. "it is an empirical fact that the vast majority of men have a disposition to feel moral emotions… it is an empirical fact that there is very substantial agreement among men in the kinds of act which call forth moral pro-emotion and in the kinds which call forth moral anti-emotion" (159-60). He also mentions generalisations about the sorts of conditions in which moral emotional dispositions are masked (161-2).

Given this, it may seem that Broad's amended Trans-Subjectivism falls foul of the objection that he raised against Hume, i.e., moral inquiry and dispute is ultimately a matter of empirical investigation and statistics gathering. As Frankena (1959: p. 560) put it: even if the Trans-Subjective analysans is analytic, "a statistical

study is involved at least at second remove".

Whether or not one finds Frankena's worry at all potent, a complicating factor is that Tran-Subjectivism may be thought of or developed as a version of the Ideal Observer Theory defended later in the twentieth century by the likes of Roderick Firth (1952), according to which moral propositions are analysable in terms of the attitudes of a suitably defined ideal agent. Indeed, Firth cites Broad's presentation of Trans-Subjectivism in Reflections as a forerunner of this view (there is also a case to be made for thinking that Hutcheson held an Ideal Observer account). If that's right, then maybe it makes it easier for Trans-Subjectivism to avoid Frankena's worry.

III.

What Broad calls 'Trans-Subjectivism' can be thought of as a species of Sentimentalist logical analysis: it analyses moral propositions partly in terms of sentimental or emotional responses. In recent years Trans-Subjectivism has been eschewed among many Sentimentalists in favour of Neo-Sentimentalism: evaluative and moral propositions are analysed in terms of *merited* or *justified* emotional responses. This arguably addresses the worry that it is not inconceivable that we might think that the emotional responses of normal agents in normal circumstances are inappropriate. A Hutchesonian Neo-Sentimentalism might claim that the proposition that *acts of promise keeping are pro tanto right* is analysable as *approbation is justified/merited in response to the contemplation of acts of promise-keeping.*

Although Neo-Sentimentalism makes reference to emotions in the analysans, one might think that its adoption rules out the view that emotions could constitute perceptions of values as Representational theorists claim (an epistemological implication). Here are two reasons: Broad along with other contemporaries[14] discussed what I'm calling 'Neo-Sentimentalism' as a version of non-naturalism. Because of this they assumed that the way in which we recognise that certain kinds of emotional response are justified or merited is via Reason or intuition. Further, on this account emotions were not thought of as representations or perceptions of moral properties. One reason for thinking this is that this would threaten to introduce circularity in to the analysis, i.e., if approbation were a perception of rightness then the Neo-Sentimentalist analysans would be *acts of promise keeping merit perceiving them as right*. But now *right* appears on both sides of the biconditional.

Regarding this, two things are worth mentioning. Firstly, we need not think of Neo-Sentimentalism as a

14. A Fitting-Attitude analysis of value (which is, strictly speaking, distinct from Neo-Sentimentalism) is endorsed by Alfred Ewing (1953).

non-naturalist view. It has contemporary defenders who deny non-naturalism (e.g. Gibbard (1990)). Even if non-naturalism necessitates a Rationalistic epistemology (something I'm unsure about) Neo-Sentimentalism need not entail this. Second, some proponents of Neo-Sentimentalism embrace the circularity described, adopting a No-Priority view (e.g. McDowell (1998)).

One might still think that Neo-Sentimentalism is incompatible with a Representational epistemology because the combination of the two views would entail that emotions could justify themselves.[15] On standard epistemological views about experience, perceptual experiences are sources of immediate justification. However, if approbation were a perception of rightness, and if Hutchesonian Neo-Sentimentalism were true, then approbation towards X would be a perception with content *approbation towards X is justified*. Subjects would thereby have immediate justification for believing that their approbation was justified. But no mental item can justify itself in this way.

In response to this, Hutchesonian Neo-Sentimentalists who are attracted to Representational Objectivism (for the purpose of readability I'll hereafter refer to this combination of views as 'Neo-Sentimentalism') might borrow a point that Broad makes in Reflections: it is generally false that "if the correct analysis of the proposition 'S is P' is 'S is p1-and-p2' then anyone who is believing the former proposition is *ipso facto* believing the latter… it is quite obvious that a number of persons who accept different and incompatible analyses of a proposition may all believe it" (163-5). Neo-Sentimentalists might simply apply this point to the case of emotional perceptual experience. Just because Neo-Sentimentalism is the correct analysis of evaluative and moral terms, this needn't entail that emotional experiences of moral properties have this content. So there need be no worry about self-justification.

However, it will likely be objected that even if this response is successful, it is still plausible that subjects who endorse the Neo-Sentimentalist analysis will be in a position to form judgments (perhaps immediately justified judgments) on the basis of experiences of approbation that do have this content. So there remains a problem of self-justification.

In response Neo-Sentimentalists should argue that they are in no worse a position than other theories which claim that experiences, e.g., sensory experiences, can immediately justify judgments. Let me explain. If, e.g., a visual experience of a red ball can immediately justify a belief that *there is a red ball* then it seems that such experiences could justify subjects in believing that their

15. For this claim see Brady 2013.

experience is *accurate*. This is because their experience could also justify them in believing that *they are having an experience of a red ball*. Together with the justified belief that *there is a red ball*, they could easily infer and come to have a justified belief that *their experience is accurate*.

What does this have to do with Neo-Sentimentalism? There is controversy as to what the relevant sense of 'justification' is in the Neo-Sentimentalist analysis. Some philosophers (e.g. Tappolet (2011)) think that it simply amounts to *accuracy*. If that were right then it is far from obvious that Neo-Sentimentalists are therefore in a worse position than those who think sensory experience can justify empirical beliefs. The only difference would be that the justification is more direct. But I'm not sure this makes a significant difference. Suppose instead that the relevant sense of 'justification' is a sui generis or distinctive emotional kind. Although there wouldn't be a strict analogy between this and, e.g., the sensory case, it is far from obvious that the problem becomes any worse (or any better) given the assumption that emotions can justify beliefs that are (or the type of experience is) in some sense required by its object, e.g., promise-keeping.

Thus, I think that the 'self-justification' argument I have considered fails to establish that Neo-Sentimentalism and a Representational Moral Sense Theory epistemology are incompatible.

REFERENCES

Audi, R. (2013), *Moral Perception*. Princeton NJ: Princeton University Press.

Brady, M. (2013), *Emotional Insight*. Oxford: Oxford University Press.

Broad, C.D. (1925), *The Mind and Its Place In Nature*. London: Kegan Paul.

_____ (1934), *Five Types of Ethical Theory*, London: Kegan Paul.

_____ (1945), 'Some Reflections on Moral-Sense Theories in Ethics', *Proceedings of the Aristotelian Society* 45: 131-166.

Chudnoff, E. (2013), *Intuition*. Oxford: Oxford University Press.

Cowan, R. (2015), 'Perceptual Intuitionism', *Philosophy and Phenomenological Research* 90: 163-193.

Dancy, J. (forthcoming) 'Intuition and Emotion', *Ethics*.

Ewing, A.C. (1953), *Ethics*. London: English Universities Press.

Filonowicz, J.D. (2008), *Fellow-Feeling and the Moral Sense*. Cambridge: Cambridge

University Press.

Firth, R. (1952), 'Ethical Absolutism and the Ideal Observer' *Philosophy and Phenomenological Research* 12 (3).

Frankena, W.K. (1955), 'Hutcheson's Moral Sense Theory', *Journal of the History of Ideas* 16: 356-75.

_____ (1959), 'Broad's Analysis of Ethical Terms' in P.A. Schipp ed. *The Philosophy of C.D. Broad*. New York: Open Court.

Gibbard, A., (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge MA: Harvard University Press.

Harman, G. (1977), *The Nature of Morality*. Princeton NJ: Princeton University Press.

Hutcheson, F. (1969a). *An Inquiry Concerning Moral Good and Evil*, originally published 1725, in D.D. Raphael ed. *British Moralists* 1650-1800. Indianapolis, IN: Hackett.

_____ (1969b). *An Essay on the Nature and Conduct of the Passions, with Illustrations Upon the Moral Sense*, originally published 1728, in D.D. Raphael ed. *British Moralists* 1650-1800. Indianapolis, IN: Hackett.

Kauppinen, A. (2014), 'A Humean Theory of Moral Intuition', *Canadian Journal of Philosophy*, 43: 360-81.

McDowell, J. (1998), *Mind, Value, and Reality*. Cambridge MA: Harvard University Press.

Prinz, J. (2004), *Gut Reactions*. New York: Oxford University Press.

Russell, B. (1912), *Problems of Philosophy*. London: Home University Library.

Siegel, S. (2010), *The Contents of Experience*. New York: Oxford University Press.

Slote, M. (2009), *Moral Sentimentalism*. Oxford: Oxford University Press.

Tappolet, C. (2011), 'Values and Emotions: Neo-Sentimentalism's Prospects' in Bagnoli, C. ed., *Morality and the Emotions*. Oxford: Oxford University Press. pp. 117-134.

_____ (2012), 'Emotions, Perceptions, and Emotional Illusions' in Calabi C. ed., *Perceptual Illusions. Philosophical and Psychological Essays*. London: Palgrave-Macmillan. pp. 207-24.

MICHAEL TANNER studied at Corpus Christi College, Cambridge, where he has been a fellow since 1961. He was for many years a lecturer in the Philosophy Faculty at Cambridge. Within philosophy he is best known for his work on aesthetics and on the history of post-Kantian European philosophy, especially Nietzsche. He has also written on literature, and is an influential music critic and expert on the works of Richard Wagner.

by Michael Tanner

I WANT IN THIS PAPER to raise some questions concerning the rôle and importance of examples in moral philosophy, and to consider very briefly some ethical controversies to see whether they can suggest, and in turn be illuminated by, answers to them. In recent years there has been an undercurrent of complaint about the examples which twentieth century Anglo-Saxon philosophers have usually taken to illustrate their analyses of ethical concepts and arguments. Thus, in the peroration to her book *Ethics Since 1900,* Mary Warnock writes, "It is impossible to predict what kind of books moral philosophers of the future will actually write. But the examples which they contain will necessarily have to be long, complicated, and realistic. I think that the days of shouting to revive the fainting man, and the days of grading apples, are over. Moral philosophy will be much more difficult, perhaps much more embarassing, to write than it has been recently, but it will be far more interesting to read."[1]

And sometimes the consequences of the kind of examples that we usually find are alleged to be very serious: "A point of method I would recommend to the corrupter of the youth would be this: concentrate on examples which are either banal: you have promised to return a book, but . . . and so on; or fantastic: what you ought to do if you had to move forward, and stepping with your right foot meant killing twenty-five fine young men, while stepping with your left foot would kill fifty drooling old ones. (Obviously the right thing to do would be to jump and polish off the lot)."[2]

Often, though by no means always, these complaints about triviality, banality, and far-fetchedness are accompanied by a rather nervous glance across the Channel: continental moral philosophers, we know, not only write theoretical treatises which we find stimulating if unintelligible, but also novels and plays. Being moral philosophers, whatever they write is moral philosophy. Hence at least one way of doing moral philosophy is to write novels and plays. This line of argument is fascinating, but unclear: if the suggestion is that these are *alternative* ways of doing the same thing, we need a lot of explanation before we can accept, or even understand, such a claim.

1. pp. 206-207.

2. G. E. M. Anscombe in *The Listener,* Feb. 14, 1957: p. 267.

It might be alleged that I am stating the existence of a widespread phenomenon without being able to offer much evidence of it: and I agree that there are not many places in which we can find this kind of complaint expressed in writing, rather than voiced. It is, admittedly, the sort of thing that people say at meetings rather than write in articles or books. That does not mean that we shouldn't take account of it; it does mean, however, that to a considerable extent I must construct a case before I examine it.

The argument would, I think, go like this: the subject-matter of moral philosophy is the whole set of concepts which are employed in making moral judgments and in justifying them. This set of concepts is large and various, and the arguments in which they occur are similarly diverse. But what we find in Anglo-Saxon philosophy is an almost exclusive concentration on what are taken to be the key ethical terms, namely, "good", "right," "ought", and "duty". While these certainly are important, it is easy to overstress the role they play, and "good", possibly the most favoured of all, doesn't really "come in very much except in the pages of books about moral philosophy."[3] The result of this is that moral philosophy has become parochial, partial, boring, and irrelevant to practice. And this tendency has been fostered by the extremely limited range of examples which we customarily find in ethical textbooks. If moral philosophers would look, or look again, at the whole set of phenomena which constitute our moral life, and at the language in which we talk about them, if they would start from, and keep more constantly in mind, the concrete manifold, they would stand a good chance of producing work that was unparochial, comprehensive, interesting, and relevant to practice.

This kind of argument is by no means to be found only in ethics, though it is here at its most insistent. Perhaps it is worth asking first why we may more easily feel sympathy with it in ethics than in epistemology. To which the answer may be that, however we *do* feel, we *should* feel the same way about epistemology too. Wittgenstein clearly felt it in a quite general way: "A main cause of philosophical disease—a one-sided diet: one nourishes one's thinking with only one kind of example."[4] And so did Austin: the view that he examines in *Sense and Sensibilia* is, he says, "a typically *scholastic* view, attributable, first to an obsession with a few particular words, the uses of which are oversimplified, not really understood or carefully studied or correctly described; and second, to an obsession with a few (and nearly always the same) half-studied 'facts'. (I say 'scholastic', but I might just as well have said 'philosophical'; over-simplification, schematization, and constant obsessive repetition of the same

3. Mary Warnock, *op. cit.,* p. 202.

4. *Philosophical Investigations,* § 593.

small range of jejune 'examples' are not only peculiar to this case, but far too common to be dismissed as an occasional weakness of philosophers.)"[5] However, I think that Austin does not substantiate his claim. Though he exposes many fallacies in the philosophers with whom he deals, it does not seem clear that they are the result of their confining themselves to the same small range of jejune "examples." What Austin primarily shows is that the examples themselves have been seriously misdescribed: and he clarifies matters by considering those same examples, and others not significantly different, more intensively, and characterizing them more accurately, than previous philosophers had done. None of his major points depends on producing further, and more varied examples. The main point here is that we do know what range of statements we want an analysis of when we are concerned with the philosophy of perception: and although Austin complains that the expression "material thing" is a piece of jargon introduced without adequate explanation, and that the examples which are customarily provided of material things don't sufficiently delimit the concept, he does not succeed in showing that any serious consequences follow from that—from a concentration on "moderate-sized specimens of dry goods."[6] And is it really true that Price "complicates matters" by adding "water" and "the earth" to *his* list of specimen material things, as Austin alleges?[7] Surely it is clear that a satisfactory analysis of "There is a chair in the next room" would serve, *mutatis mutandis,* for "there is a river beyond the next field", or for statements involving flames, rainbows, or planets.   So it remains true that the gravamen of Austin's complaint is that the examples which *are* considered are incompletely or inaccurately characterized, and that generalizations on the basis of these perfunctory characterizations are made far too hastily. We may be bored stiff by elliptical pennies; but would it help much to change to trapezoid books?

Much the same may be said, I think, about the use of examples in all the central epistemological problems. What we find, time and again, is not so much an unwarranted assimilation of one particular case to another, as a refusal or inability to attend properly to any *one* particular case. The assimilation may be unwarranted, but when the *assimilans,* if I may coin the term, has in any case been given far too short shrift, it is impossible to say what has gone wrong in dealing with the *assimilandum.*

If this is correct, then we may have come across a major difference between the subject-matter of epistemology and that of ethics. For while, when we do epistemology, we are dealing, in Strawson's words, with the "massive central core of human thinking which has no history—or none recorded in histories of thought—cat-

5. P. 3.

6. *Op. cit.,* p. 8.

7. *Ibid.* fn. to p. 8.

egories and concepts which, in their most fundamental character, change not at all,"[8] this cannot be said of the categories and concepts which are central to morality, which have all too much of a history. I am not going on to argue that this shows a fundamental difference between "factual" and "evaluative" judgments and concepts, though others might. I am only pointing out that the first-level material with which the moral philosopher has to deal is deeply controversial in a way that is not the case with the material of epistemology. A consequence of this is that an enormous amount hangs on what examples we pick in moral philosophy. To make clearer that and how this is so, I propose a distinction between "paradigm" and "example", which will enable me to characterize the complaints which I am considering as claiming that moral philosophers are prone to confuse the two, and to imagine that they are producing paradigm when in fact they are only producing examples. I am not going to offer precise definitions of the terms, but I think the distinction which I have in mind is sufficiently indicated by saying that x is a paradigm of an o if the salient features of x are salient features of anything which is an o, and if there are no properties which it is essential for an o to possess which x does not possess; while x is an example of an o if it possesses properties sufficient to count as an o, while also possessing salient features which do not so count, and if it is not essential for something's being an o that it should have all or even a large proportion of the properties of x. Or, in terms of necessary and sufficient conditions, x is a paradigm of an o if the properties of x which are the sufficient conditions for its being an o are also necessary conditions for anything's being an o. But if the sufficient conditions for being an o contain sub-sets such that the possession of one sub-set is sufficient for y's being an o when y has no or few properties in common with the properties of x which are sufficient condition for x's being an o, then x is only an example of an o.

This distinction is intended to apply primarily to statements, and the specific statements offered as specimens for analysis in philosophy, which is why I stipulated that the salient features of x should be the salient features of anything that is an o; applied to objects, *etc.,* this condition would not be desirable. That it is necessary in this context I shall demonstrate in due course.

Using these terms, then, we may paraphrase "a one-sided diet" as "treating examples as if they were paradigms", and the complaint which we are considering against contemporary moral philosophers would similarly be that what are only examples of moral judgment, *etc.,* are treated as paradigms: so that the salient features which a specimen moral judgment possesses are erect-

8. *Individuals,* p. 10.

ed into necessary conditions of anything's being a moral judgment. The result is that if two philosophers take as paradigms of moral judgments cases which, while they possess properties sufficient for them to count as moral judgments, do not share certain salient features, and if these salient features are erected into necessary conditions of a judgment's being a moral judgment, then we are offered two contradictory accounts of morality, with the opposing philosophers each pointing out that the other's account is hopelessly wrong, since the general analysis which is offered, derived from his paradigm, fails to take account of the features which are brought so clearly to mind by consideration of the other one.

Presented as abstractly as this, my description sounds like a parody of philosophical conflict. But it is very nearly what has happened, as I shall show by taking several examples. My first is from Stuart Hampshire's article "Logic and Appreciation".[9] In that article Hampshire makes a sharp distinction between ethical and aesthetic judgments, the upshot of which is that there are not, or should not be, any aesthetic evaluations. Now it might be said that the trouble, or one of the troubles, with Hampshire's article is not that his examples are limited, but that he does not take any examples at all. But while it is true that no specific cases of ethical or aesthetic judgments are mentioned, it is perfectly clear what kind he has in mind. His paradigm of a moral judgment would be "You ought to do *x*", when one is telling someone what course of action to pursue in a specific situation; or "You ought always to do action of type *x*"—suitably qualified, no doubt, to take account of difficult cases when one is laying down a general moral rule. That these kinds of judgment *are* what Hampshire has in mind as paradigms, in my sense of that term, comes out very clearly if we recall his general line of argument, which is that the point of moral judgments is to get people to act in certain ways in order that they may solve the problems which confront them. All moral judgments involve recommendations, and the characteristic activity of the moralist[10] is such that "anyone who moralizes necessarily generalizes: he 'draws a moral', in giving grounds of choice, he subsumes particular cases under a general rule" *(ibid.,* p. 164). In contrast to this, "judgment of a work of art does not involve a choice", and "a critical judgment makes no recommendation" *(loc. cit)*. For the purpose of the critic is "to look here, at previously this unique object, to see it as unique and unrepeatable" *(ibid.,* p. 165), and so "the spectator-critic in any of the arts needs gifts precisely the opposite of the moralist's: he needs to suspend his natural sense of purpose and significance" (p. 166).

I am not going to criticise Hampshire's argument,

9.   To be found reprinted in *Aesthetics and Language* ed. W. Elton; all my page references are to this book.

10. By "a moralist" Hampshire does not mean more than "one who makes moral judgments": we are all moralists. This use is common practice now and significant.

though it seems to me to be wholly mistaken, and to involve manifold and manifest confusions. Taking for granted that it is wrong, I want merely to ask how it has gone so wrong. We are offered a general characterization of moral judgments, aesthetic judgments are shown to be markedly unlike them in all essential respects, and so we are to conclude that to evaluate works of art in an analogous way to that in which we evaluate actions morally will involve our "making unnecessary choices and laying down principles of exclusion, as a moralist must" (p. 168). The simplest and most effective way to reply to this whole argument, if one were doing that, would be merely to give a specimen list of moral judgments, from which one could immediately see how few of them fitted in with Hampshire's characterization. It might then be said that his argument was nothing more than an exemplification of the confusion between generality and universality: the harm that this confusion has wrought in the history of ethics has still not been sufficiently realized. But that confusion itself is a result of bearing in mind only general moral commonplaces. It is only if one is thinking of them as paradigms of moral judgments that one is able to say, "Anyone who moralizes necessarily generalizes" and mean by that more than that anyone who makes a judgment on a certain situation is committed to making the same judgment on any exactly, or relevantly, similar situation. If one does *not* mean more than that, one is merely stressing the rationality of moral judgments, in the sense that one is demanding that they be consistent. And from the demand for consistency in morality nothing follows as to the difference between moral discourse and other forms of rational discourse, of which aesthetic discourse may still be a member; quite the contrary—the universalizability-criterion assimilates, it does not differentiate.

What we find in Hampshire, then, is both reliance on a far too limited range of (implicit) examples, and extremely inadequate description of those that he considers to be paradigms. The first shortcoming, indeed, may be said to foster the second. If he had borne in mind judgments like "This act was very courageous", or "He is a noble person" or the kind of advice that one would give to someone in a highly complex situation, he would not have been able to make many of the statements he does. He would not have seized, for example, on "Some force of recommendation" as being a salient feature of "every moral judgment", without at the very least explaining far more fully what he meant; for the more we consider examples other than those of the type "You ought to do *x*", the less clear it becomes that much is being said in the claim that all moral judgments have recommendatory force (or, for that matter, that they have commendatory force), and so the less clear, equally, it becomes that

it is on account of *that* feature that the judgments which unquestionably and straightforwardly possess it are moral judgments. Again, returning to the generality of some moral judgments: this leads Hampshire to contrast the activity of the moralist with that of the critic, looking "here, at precisely this unique object".   But does the generality of some moral judgments (all, according to Hampshire), mean that the moralist is not required to look here, at precisely this situation? What Hampshire, like very many other philosophers, has in mind, are those situations when there is one feature which is (almost) always of overriding importance, such as the taking of human life or the violation of property or the relief of pain. But even here, most people would admit, each situation must be considered on its merits, and the implication that the moralist is somehow excused, since his concern is with assimilation of one situation to another, from paying close attention to any given situation, is an odd one. The picture that lies behind all this, of our approaching moral problems armed with a set of general principles, and classifying the problems that we meet so that one or another, or sometimes more than one, of the principles apply to them, is as profoundly mistaken as an ethical view could be. It has, of course, very complex roots, like any widespread and long-held position; but I think that it is no exaggeration to say that if the philosophers who endorse it had reminded themselves of the variety of ethical judgments, they would have thought harder about continuing in their allegiance to it.

Similarly, the view that there are not or should not be any aesthetic evaluation, which is held not only by Hampshire, could only be maintained by someone who, in referring to "works of art", was thinking only of those works which are generally agreed to be of major stature, so that evaluation or re-evaluation of them comes to be, or seem, pointless, and so, perhaps, does comparative judgment of them. But as soon as one remembers that Swinburne produced works of art as well as Shakespeare, or Tretchikoff as well as Titian, it becomes immediately clear that evaluation is as imperative in art as it is in morality. And if, in reply to this, someone were to ask, "But in accordance with which general principles is art to be evaluated?" in the tone of voice of one who has worked through many treatises on aesthetics without finding any satisfaction, the best answer would be a *tu quoque* about morality. Not that general principles aren't to be had there, but perhaps a fairer estimate and a deeper understanding of their role and importance, at this stage, would come from studying aesthetic argumentation, and then comparing moral argumentation with it, than from the reverse process, which we are all used to.

A series of confusion interestingly similar to Hamp-

shire's is to be found in Bernard Mayo's book *Ethics and the Moral Life*.[11] We find there a distinction between moral relationships and personal relationships, the latter, it is said, being amoral; and the arguments deployed for this conclusion parallel Hampshire's very closely. In moral relationships, Mayo says, we are concerned with the features which one has in common with another, while in a personal relationship we are concerned only with becoming aware of the uniqueness of the other person. Once again, in studying this contrast, we are aware of the way that both parties to it have been falsified, and we see how much of the trouble comes from thinking of morality as consisting entirely of rule-of-thumb principles. It is not an accident that, as well as producing his curious account of personal relationships, Mayo confuses generalisability with universalisability: for if one does that, and recognizes that moral judgments are universalisable, the rest is more or less certain to follow. The exaggerated importance assigned to the universalisability of moral judgments in recent years results partly from this confusion, partly from the excitement of finding that it enables one to combine subjectivism in ethics with the possibility of moral argument, and partly from recognizing its close connexion with that potent moral weapon, "But suppose he was in your situation, and you were in his?" Once again, a larger range of examples both within ethics and, for purposes of comparison, outside it, would show that this is not a logical peculiarity of ethics, but that since morality involves, often, acting against one's desires, it is more often here than elsewhere that people are highly inclined to act without regard for consistency: but this is a purely contingent matter.

What adds to the confusion about generality in morals is that those general maxims which spring so readily to mind are, in some sense, basic: perhaps the best way of stressing this and also avoiding being hypnotised by it is to adopt Strawson's formulation, that they are "of the first importance as a condition of everything that matters, but only as a condition of everything that matters, not as something that matters in itself."[12] Judgments on situations which involve these matters can usually be made without explicit reference to ends, but this does not mean that the judgments do not in fact concern ends. Not to realize this lands us in the second major, and historically enormously more important, confusion resulting from this particular one-sided diet: which is, roughly speaking, deontology. I am aware that it would be absurd to dismiss the whole set of views which is designated by this term as a result of nothing more than this. Other grounds for being a deontologist exist, and some of them may be more respectable. Nonetheless, a good deal of the impressiveness of deontology has come from the important place that judgments involving no

11. See Chapter X, Sections 8 and 9.

12. "Social Morality and Individual Ideal", in *Philosophy*, 1961: p. 5.

reference to ends have in moral discussion. A deontologist will feel that there is no need to justify such judgments; while a teleologist will say that the ends which they serve are so invariable, and our attitude towards them so firm, that it is hardly ever necessary to make explicit reference to them.

In reaction from taking these general rules as paradigms of moral judgment, and the situations to which they apply as paradigmatic moral situations, some philosophers present us with situations so complex, involving so many counter-balancing factors, that it is clearly impossible, in deciding what ought to be done, to appeal to any pre-existing rule, general or otherwise. Contemplation of such situations leads to talk about creating one's values, choosing in the void, and so on. The *locus classicus* here is Sartre's example of the boy who has to choose between staying at home to look after his mother and going to fight for his country; though this situation is in fact less complex and no more agonising than innumerable other dilemmas that people are incessantly faced with. Philosophers who work from this example tend to produce accounts of moral concepts, decisions, *etc.,* which are in sharp conflict with those we have already considered. What, at last, may now be possible is a recognition that neither the kind of case that Ross presents us with, nor that which Sartre offers, is to be taken as the paradigm of a moral judgment or situation. As far as complexity is concerned, and thus the applicability or not of general principles which contain no reference to ends, there is no such thing as a paradigmatic moral situation. Some are extremely straightforward, only needing classification for us to know how we should judge or act; some are extraordinarily elaborate, so that precedents and rules will only take us part of the way: and most, at least under normal circumstances, come somewhere in between. It should, consequently, be clear that an adequate moral philosophy will be able to cope with and account for cases which come under any of these heads; and so any statement we make by way of general characterization of moral judgments and concepts must be tested against a wide range of examples.

The important word here is *range.* Those who complain about examples have not, as I observed earlier, made clear precisely what their grounds of objection are. I hope to have indicated that it is not triviality, triteness, or familiarity that are the trouble: we don't, in general, want examples which will engage our moral feelings rather than our meta-moral concerns: though by too much repetition we may become unable to concentrate properly on what characteristics the examples actually possess. At this stage we have not found that there is any *desideratum* other than variety.

Variety, however, may be of several different kinds. Do we merely want to see "ought" used in a situation to which a basic general rule applies, and in a Sartre-type situation, and in reference to the actions of an historical personage, and in non-moral contexts of several kinds, and so on? And similarly, do we want to study the way "good" functions when applied to actions, objects, ends of action, people, and organizations? Or is it important to draw on the full range of moral predicates, and study the behaviour of more than the most general, if not the most common, ones? These questions take up Mary Warnock's point, mentioned earlier, that it is time that moral philosophers concentrated on more specific, and (it is claimed), more frequently used, moral terms, and is in the spirit of Austin's "If only we could forget for a while about the beautiful and get down instead to the dainty and the dumpy."[13] There is a good deal more to this suggestion than appears at first sight, and for most of the remainder of this paper I shall pursue some of its implications, by considering it in relation to the controversy between Hare and Mrs. Foot on whether a ("pure") description of an action can entail an evaluation of it. This matter, which is taken by many contemporary moral philosophers to be the most fundamental of their problems, raises very many issues, nearly all of which I shall ignore. What I want to consider in a little detail are the methods which Mrs. Foot employs for arguing against the fashionable relating of "goodness" to "choice". Hare's contention is that, though much more moral argumentation is possible than is often thought by subjectivist-inclined philosophers, when it comes to the push we *choose* our criteria of goodness. And this is taken to be an alternative formulation of the thesis that no evaluation is entailed by a description. Now, pretending that we are clear about the meanings of these terms, the way in which we are most likely to be convinced of the truth of this claim is thinking how the most general term of commendation in English is used. I say "This is a good film because it's amusing, takes you out of yourself, *etc.*" while you say that it's a bad film because it has no relevance to contemporary life, no serious theme which it can develop, and so forth. No one denies any longer that there may be argument about these criteria themselves as well as about their application. But if we go on disagreeing, in the end we must just agree to differ, and neither of us can be said to be right or wrong. And, so the argument is taken to imply ("*No* description entails an evaluation"), what goes for the general goes for the more particular, too. One studies "good" merely because it *is* the most general value-term, and thus may be taken as a paradigm of all value-terms.

Mrs. Foot suggests, on the other hand, at least in some of her work, that we might start by looking at

13. *Philosophical Papers,* p. 131.

more specific value-terms than "good", to see how they work—the equivalent in ethics of the dainty and the dumpy. She takes "rude" as a specimen value-term, and points out (here I am summarising the argument of part of her article on "Moral Arguments")[14] that "rude" expresses disapproval, is used when an action is to be discouraged, and so on. She then shows that we cannot choose our descriptive criteria for calling an action rude—only that behaviour is rude which gives, or tends to give, offence; and only certain types of behaviour *do* tend to give offence. So from a description of how some one behaved we can deduce that he was rude, and hence this his action is to be condemned.

14. *Mind*, 1958. See particularly pp. 507-9.

What does this argument show, supposing that we accept it? Chiefly, I am inclined to say, that we must not forget for a while about the right and the good and get down instead to the rude and the grateful. For it may be true that in the case of relatively specific value-terms, a description entails an evaluation, while for more general terms this does not hold. And this possibility may be realized without invoking the concept of "pure value-terms". I mean that we might well drop certain specific value-terms altogether, or undertake a heavy programme of redefinition; but we could not, as long as we continued to be men, abandon the use of "good" and "right", or go "beyond good and evil" (not that Nietzsche meant what that title suggests). Mrs. Foot admits, and Hare insists,[15] that in the case of such specific value-terms as "rude", it would be possible for us to abandon them, but the consequencies of this have remained unclear. That such things not only might, but do, happen, can be seen from the obsolescence, at least in most circles, of, *e.g.,* "chaste" or "gentleman", evaluative terms for which one clearly cannot choose one's descriptive criteria. But the questions remain as to how much the fact that there are such terms represents a triumph of naturalism, and as to whether their actual or possible waxing and waning in a society represents a successful come-back by anti-naturalism (I mean Hare's kind, not Moore's). It seems clear that the answer to the first question is that naturalism only *has* triumphed by this demonstration if it can be extended to the most general moral terms. Mrs. Foot edges towards saying that it can in the article from which I have already quoted, when she says, "(A man) can no more choose what shall count as a benefit than he could have chosen what counted as harm,"[16] and in a later paper when she says "My thesis is not, of course, that the criteria for the goodness of each and every kind of thing are determined in the same way as they are determined for such things as knives, but rather that they always are determined, and not a matter for decision."[17] But when she makes these claims, we once more feel a crying need for more examples. Let

15. See *Freedom and Reason*, p. 189.

16. P. 511.

17. *Proc. Arist. Soc., Supp. Vol.,* 1961, p. 67.

us agree that in the case of some behaviour, a description entails an evaluation; that in the case of terms which are defined functionally, such as "knife", we cannot choose what is to count as a good knife, though we may be able to choose what are good knives for our purposes, or to disapprove of any use to which a knife could be put; and that we do not have any, or much, choice as to whether it is correct to say that a plant is flourishing. We are still a long way from having agreed that we can't choose our criteria, or some of them, for an action's being condemnable, for a man's being good, or for his flourishing. To take the case of the criteria for something's being a good member of its kind: granted that the kind is defined, like "knife", both functionally and in terms of a set of observable properties,[18] then we have at least the following four classes:

1). Cases where a certain kind of thing has a given function, or answers to a given purpose.

2). Cases where more than one kind of thing has a given function, or answers to a given purpose.

3). Cases where a certain kind of thing has several functions, or answers to several purposes, which may be such that excellence in fulfilling one of them reduces its efficiency in fulfilling some others.

4). Cases where several kinds of things have several functions or serve several purposes, some of which are shared by the different kinds of things, and some not.

Now though it may be true that (1)- and (2)- type cases answer to what Mrs. Foot says about the determination of the criteria for goodness, it is by no means clear that (3)- and (4)- type cases do. And the term "man", which would fall under type (3), *i.e.,* men have many different functions, purposes, *etc.,* certainly does not seem to come within the scheme she presents. Would Mrs. Foot maintain that there is, or ever will be, a definition of "man", or a definitive account of the nature of men, from which it is or will be possible to read off the correctness or otherwise of my attribution of goodness to a particular man? Another way of putting this is that whenever we are confronted with a concept to which the family-resemblance treatment is appropriate, not only as regards the observable properties which the object designates (games played with a ball, by twenty-two men, on a field of certain dimension, and games played with a pack of cards by two people, *etc.*), but also as regards the function which it fulfils or the purposes it serves (to provide *corpora sana* for *mentes sani* to inhabit, to provide recreation, to provide a living for the players, to pass the time, to excite spectators, *etc.*), then there may arise legitimate differences as to which things instantiate

18. "Knife: Blade with sharpened longitudinal end fixed in handle either rigidly, or with hinge, used as cutting instrument or weapon." *O.E.D.*

the concept better than others. What such a dispute may finally issue in, in the case of anything except a dispute about "man", is a discussion, with a view to comparative evaluation of, sets of needs, interests, desires, and preferences. A dispute about whether one thing is better than another, which is unsettlable, when we have all the necessary information about the things themselves, must become a dispute about the purposes which those things serve; a dispute, in other words, about ends, and thus about the nature of the beings who have those ends. And neither Mrs. Foot nor anyone else has done much to show that her account of knives and plants works there.

It is in the light of this that we must reconsider the significance of *(a)* the fact that from the definition of a concept one can deduce what would be a good object of the kind the concept designated, and *(b)* the fact that there are many concepts which we can choose not to employ, and many purposes which we can choose not to implement. MacIntyre, in his review of *Freedom and Reason,* says "How the undeniable fact that I may choose not to use a given concept could alter the character of that concept I cannot see,"[19] in reply to Hare's point about "rude", *etc.* Hare may have been obscure on this point, but I think he is basically correct. He is arguing on a broader front than MacIntyre suggests, and his argument uncovers the fact that Mrs. Foot appears to be establishing more than she actually is. The disagreement, I hardly need repeat, is about the relationship of description to evaluation. Hare claims that there are no entailments; Mrs. Foot that there are. Hare attempts to establish his case by analysis of "good", Mrs. Foot by analysis of "rude". But is there a real disagreement here? All that Mrs. Foot succeeds in showing is that, given the meaning of certain evaluative terms, and granted that we are prepared to employ them at all, then we are not entitled to apply or withhold them in accordance with our own chosen criteria. But Hare's point is prior to this: it is about our preparedness to make certain evaluations in the first place; or about, if you like, what our moral vocabulary is to be. Mrs. Foot's argument only operates, given a moral vocabulary. But implicit in her argument, and occasionally explicit too, is the much more important and dubious claim that we have no choice as to what our moral vocabulary is to be, because we have no choice as to what to account as harms and benefits. The example Mrs. Foot takes here is that of torture.[20] Once again, this is a one-sided diet. What of choosing between one career and another? Or the merits of a short life and a happy one versus those of tame longevity? We may, in many situations, choose what to count as harms and benefits. What is the nature of the supposed impossibility here—logical or psychological?

19. In *Philosophical Books,* May 1963, p. 7.

20. *Mind,* 1958, p. 511.

I have raised questions here, rather than answered them. My main point has been that by inadequate selection of examples a genuine confrontation between naturalism and anti-naturalism has been largely avoided. We should now be able to see what is wrong with the spirit of Austin's plea about the beautiful and the dumpy: if we are going to give a comprehensive account of the logic of moral or aesthetic discourse it will be necessary to consider both the most general and the more specific value-terms: and philosophers have been right to regard consideration of the former as their primary task. What I think is now most important is an attempt to map out the relations between general and specific value-terms. For, to return to Mrs. Warnock's point, there *is* something very artificial about taking examples of commonplace actions which have goodness predicated of them. The most general terms, as I have suggested, are chiefly employed when we are confronted with a fairly basic challenge to our moral scheme, and they should be studied in that context. In other words, we need to study the structure of a whole system of morality in order to understand the postulated relationships between fact and value, and basic and derived values, which are to be found in it, rather than studying isolated statements or fragmentary arguments which might occur within several quite different total moral outlooks. And if that is what is intended by the complaints I have mentioned, then they are justified.

THICK DESCRIPTION REVISITED: TANNER ON THICK CONCEPTS AND PERSPECTIVALISM IN VALUE PHILOSOPHY

Commentary by Anna Bergqvist

ANNA BERGQVIST is Lecturer in Philosophy at Manchester Metropolitan University. Her principal research interests are aesthetics and moral philosophy. She has published on aesthetic and moral particularism, thick evaluative concepts, moral perception, narrative ethics, art interpretation, philosophy and museums and selected issues in philosophy of language and mind. She is co-editor of *Philosophy and Museums: Ethics, Aesthetics and Ontology* (Cambridge University Press, 2016) and *Evaluative Perception* (under contract with Oxford University Press). She also works on the intersection between value philosophy and philosophy of medicine, currently preparing a monograph on particularism in medical epistemology.

THICK DESCRIPTION REVISITED: TANNER ON THICK CONCEPTS AND PERSPECTIVALISM IN VALUE PHILOSOPHY

Commentary by Anna Bergqvist

I. INTRODUCTION

AS IS WELL KNOWN G.E. MOORE HELD that the thin moral property of intrinsic goodness is neither reducible to, nor constituted by, natural properties but that it supervenes or is determined by natural properties, and that we know which things are intrinsically good by means of intuition. To many philosophers, R.M Hare and Bernard Williams included (who both hold that thin concepts are not 'world-guided'[1]) this is too extravagant. They find it doubtful whether any scientifically respectable view of the world can allow for properties other than natural ones. Hare sought to make progress with the familiar qualms about Moore's non-naturalism about thin evaluative concepts by drawing a distinction between descriptive and evaluative predicates. Philippa Foot, by contrast, sought to make progress by reversing the order of explanation or analysis between general and specific value-terms. She argues that thin evaluative concepts should be understood in terms of substantive value-terms, the thick ones, where the latter are seen as inherently evaluative concepts that, if we want to say so, pick out "first-order" moral properties. In his remarkable 1964 article 'Examples in Moral Philosophy', Michael Tanner questions the terms upon which the argument between Hare and Foot have been premised in a way that calls forth another category that is precluded by the traditional dichotomy between fact and value, between objective and subjective.

Hare holds that there are no rules governing what can count as a thin moral concept because, at the descriptive level, there is nothing in common between all the things that we call good; no set of descriptive properties provides sufficient conditions for the use of thin evaluative terms such as 'good' or 'bad'. For both Hare and Stuart Hampshire, what unifies assertions of thin deontic and evaluative concepts as moral judgements is instead to be found in the prescriptive or 'commendatory' function of (thin) moral terms; they express our pro-attitudes towards certain actions.[2] To assent to an imperative, in turn, is to prescribe action, to tell oneself and others to do the corresponding action. However, this does not mean that assertions of goodness are merely non-cognitive expressions of approval. On

---

1. Bernard Williams (1979) maintains that thick evaluative concepts are "world-guided", in as much as the thoughts and judgements expressed by utterances involving terms such as 'elegant', 'garish', 'integrity' are candidates for truth and falsity. At the same time thick evaluative concepts are also held to be "action-guiding", in the sense that, as Williams puts it, 'they are characteristically related to reasons for action. If a concept of this kind applies, this often provides someone with a reason for action' (Williams 1979: 140.) William's caveat about the action-guidance or practicality of thick evaluative concepts is arguably due to his reasons internalism: S has a reason to only if there is a "sound deliberative route" from S's "actual motivational set" M to (intention to) do the action. On this reading, thick evaluative concepts provide reasons only for those who endorse it (the value it may be used to ascribe) as part of one's "insider" evaluative outlook.

2. Here is how Hare (1972) expresses his position in relation to Geach's (1956) attack on the descriptive-evaluative distinction also for the most general term such as 'good' (which Geach holds is always attributive, never predicative, because one can never know what it is for an object to be good without knowing what kind of object it is): "I maintain that the meaning which is common to all the instances of the word's use cannot be descriptive and that this common meaning is to be sought in the evaluative (commendatory) function of the word." (Hare 1972: 33).

Hare's analysis, the content of judgements involving thin moral terms is found, not in their extension (which is held to be empty), but in the functional role they play in expressing our belief about the desirability of doing certain actions and not others. Foot's attack on Hare (in her 1961 paper 'Goodness and Choice' and also in her 1958 piece 'Moral Beliefs') is that a judgement cannot be identified as a moral judgement simply on the basis of formal characteristics such as universalisability and prescriptivity. Instead, she holds goodness to be tied to human flourishing. Focussing on substantive thick concepts such as *rude* rather than the thin, Foot argues that what is common is simply that all good things are 'of the kind to perform their function well' (1961: 68-59); as she argues in her (1972), moral evaluations are "hypothetical" in the sense that they serve an end – human flourishing – and will not be considered as reason-giving by those who do not share this end.

In discussing the methods that Foot employs against Hare in her attack on the separation of descriptive and evaluative judgement, 'the fashionable relating of "goodness" to "choice"' (Tanner 1964: 195), Tanner aims to elucidate just why the disputants have gone wrong, which is so much more satisfying than the simple demonstration that they are wrong. His central claim is that the search for some paradigmatic feature (or meaning-rule) of moral judgement as such should never have been begun since its outcome is irrelevant to the question as to whether there are such rules. The explanation is indirect, and proceeds via the positive suggestion in response to Hampshire (who argues that there are no rational evaluations in aesthetics because there are no general aesthetic principles) that moral argument be modelled on aesthetic evaluation, rather than the other way round. Tanner writes:

> …if someone were to ask, "But in accordance to which general principles is art to be evaluated?" in the tone of voice of one who has worked through many treatises on aesthetics without finding any satisfaction, the best answer would be *tu quoque* about morality. Not that general principles aren't to be had there, but perhaps a fairer estimate and a deeper understanding of their role and importance, at this stage, would come from studying aesthetic argumentation, and then comparing moral argumentation with it, than from the reverse process, which we are all used to (Tanner 1964: 92).

Tanner does not elaborate on the suggestion that moral argumentation be viewed through the lens of aesthetic argumentation in his 'Examples in Moral Philosophy' essay. But both the wider context of the article as a whole and the philosophical context of the time (notably the arrival of Frank Sibley's (1959) seminal work 'Aes-

thetic Concepts' a few years earlier), suggest that what is needed is precisely a re-assessment of what rational responsiveness to reasons might, in general, be taken to be in value philosophy.

In what follows I will discuss the importance of Tanner's work for the general problem of moral relevance, and the significance of the first person in ethics.

II. UNIVERSALISABILITY AND THE PROBLEM OF RELEVANCE

Hare's (1965) disagreement with neo-Aristotelianism is complex but the feature that Tanner singles out as the most fundamental aspect in the dialectic between Foot's naturalism and Hare's universal prescriptivism ("anti-naturalism", but not of Moore's sort) is Hare's position that a "conceptual apparatus" is something that one adopts, and that adopting such an apparatus is distinguishable in principle from adopting a *moral* view, thus construed as a system of moral principles (Hare 1965: 187). As R. M. Beardsmore (1969) notes, Hare's view of morality involves a Kantian-like notion of universalisability applied to some prescriptive standard that we hold in a way that allows the speaker to *choose* her own standards, so long as we are prepared to hold it for everyone in principle.[3] Such universalised standards serve as a basis for prescriptive statements of the form "x is good" (translated as "do or choose x"). Focussing on Hampshire (1967), who held a similar view of moral judgement but rejected it for aesthetics, Tanner argues that this picture of morality involves a confused assimilation of generality and universality. He writes:

> It is only if one is thinking of [general moral commonplaces] as paradigms of moral judgements that one is able to say, "Anyone who moralizes necessarily generalizes" and mean by that more than that anyone who makes a judgement on a certain situation is committed to making the same judgement on any exactly, or relevantly, similar situation. If one does *not* mean more than that, one is merely stressing the rationality of moral judgement, in the sense that one is demanding that they be consistent. And from the demand for consistency in morality nothing follows as to the difference between moral discourse and other forms of rational discourse, of which aesthetic discourse may be a member; quite the contrary – the universalisability-criterion assimilates, it does not differentiate (Tanner 1964: 191).

Universalisability, the move from the particular case to all cases that are similar in relevant respects, needs to be distinguished from generality because the latter is thought to be a degree concept. Don Loeb (1996) argues that generality, the move from the particular case to the broad sort that *includes* the particular case, is a degree

3. Hare's use of the practical syllogism differs from that of Kant because, unlike Kant's Categorical Imperative, we are not constrained by what abstract reason allows in selecting our standards on Hare's analysis.

concept because it involves '…a measure of the relative range of application of a moral principle' (1996: 80-81). Universality, on the other hand, is different, because it does not involve reference to the particular circumstances from which it emerges (such as the agent making the judgement); one can hold a principle to be universal ' … if it can be stated without the use of any proper names or indexicals' (Loeb 1996: 81). This does also seem to be Tanner's position. In discussing Bernard Mayo's (1958) distinction between moral relationships and personal relationships, for instance, he argues that judgements on personal relationships can usually be made without explicit reference to the "uniqueness of the other person" (Tanner 1964: 193), but this does not mean that the judgements do not in fact concern personal relationships. Hence, there is no good reason to say that personal relationships fall outside the scope of morality.[4]

In questioning the default assumption that morality must depend on general principles, Tanner's position could be read as an early formulation of moral particularism, although this reading is not mandatory; as noted by Julia Driver (2012) and many others, everyone can agree that a feature's reason-giving force depends upon context. The key question is *how* context enters into the equation. One option is to adopt a standard contextualist view and say that a concept's *standards of application* can vary depending on the circumstances, where the role of context is to provide an epistemic filter whereby some, but not all, possibilities can be properly ignored. This is a standard way in which the problem of relevance has been understood (see e.g. Dancy 2004). What more needs to be said?

Well, one thing that needs to be said is that it is notoriously difficult to articulate precisely what makes a possibility sufficiently remote for it to be properly ignored (or not). The reason is that such judgements are typically made against a background of presuppositions about what is constant between circumstances in which the relation of normative support between, e.g., chastity and goodness does not hold and situations in which it does hold. We rely on background assumptions all the time in navigating the world. It may be true that if you strike a dry, well-made match, it will light. As with other generics and "for the most part" generalisations, the claim that this would be so is not rendered false by the fact that if you remove the oxygen, then the struck match will not light. Tanner's point, I think, is that moral relevance is context-dependent and that its context-dependency affects notions to which it is conceptually linked like that of criterial status in ethics and aesthetics, thus understood factively as yielding knowledge. Others have made similar claims (McDowell (1998), Dretske (1971),

4. For further discussion and defence of this claim see, e.g.. Driver (2003) and De Gaynesford (2010).

Travis (2005)). So what are the implications of this for the problem of moral argument with which we started?

III. TANNER'S METHOD

Earlier I suggested that, in recommending that moral argument be understood by comparison with aesthetic evaluation, Tanner urges us to move beyond a certain view of what an argument *must* be like for it to count as rational. Tanner's conception of the wider possibilities for philosophical argument on the subject-matter of ethics can, I think, be brought into sharper focus by comparison with Wittgenstein's idea of a 'perspicuous representation' as being a key aspect of the task of philosophy as he sees it: offering a model of comparison that 'earmarks the form of account we give, the way we look at things' (*PI* §122) in order to achieve a 'clear view' of that which is troubling us (*PI* §133). However this does not mean that there is some single philosophical method through which this is achieved. On the contrary, Wittgenstein presents the philosopher with an *open-ended* range of conceptual tools and techniques that can be used in a variety of different ways, including (but not limited to): offering 'objects of comparison' and presenting 'alternative pictures'; pointing out particular 'family resemblances' and 'neglected aspects' of our language; grammatical analysis of our use of language in practice, and so on. The real task at hand is to discern which method is the most *pointful* in each context of critical appraisal for attaining clarity and revealing meaning – to which "*whatever it takes*" would be the only answer to give in the abstract.[5]

Now, in terms of what (what we may think of as) Wittgenstein's method looks like in practice, one is reminded of Sibley's (1983) notion of a 'perceptual proof' in aesthetic evaluation.[6] The focus of Sibley's discussion is Michael Scriven's (1966) scepticism about what he calls the 'independence requirement' for aesthetic evaluation. The independence requirement is a demand on rational (aesthetic) thought that 'we must be able to know the reason or reasons for a conclusion without first having to know the conclusion; otherwise we can never get the reason as a means to the conclusion' (Sibley 1983/2001: 115). In its strongest form, the independence requirement demands that reasons must be *logically prior* to aesthetic verdicts (as opposed to temporally prior in perception). Like Wittgenstein before him, Sibley does not attempt a refutation of the sceptic by way of showing how the independence requirement could be met. Instead he effectively uses the strategy of offering a 'perspicuous representation' of art criticism by pointing to the way it is actually practised to show that aesthetic evaluations stand in no need of external validation. He

5. The meaning of the notion a 'perspicuous representation' is controversial within Wittgenstein scholarship. Read and Hutchinson argue that the notion of a perspicuous representation is not to be understood as a way of seeing things and there cannot be multiple perspicuous ways of seeing the rules of 'our grammar'; any difference we might perceive between multiple perspicuous representations of an area of our grammar is merely a difference in how they are selected and arranged, something that *can* vary depending on the purpose of the investigation. (In this respect, perspicuous representations are seen as 'additive', in as much as we can combine multiple perspicuous representations of a thing's parts in order to gain a perspicuous representation of the whole.) Whether or not this is the best representation of Wittgenstein's position falls beyond the scope of this paper. I am inclined to agree with Gregory Currie (1993) (who in turn follows John McDowell) that a *representation* (as used in ordinary contexts) that transcends any point of view seems incoherent, but I cannot argue for this claim here. For further discussion see, e.g., Moore (1997), Baker (2006), Read and Hutchinson (2008).

6. Sibley first introduced the notion of a 'perceptual proof' in his seminal (1959).

writes:

> How a critic manages by what he says and does to bring people to see aesthetic qualities they have missed has frequently puzzled writers. But there is no real reason for mystification. […] What mainly is required is a detailed description of the sorts of things critics in fact do and say, for this is what succeeds if anything does; the critic may make similes and comparisons, describe the work in appropriate metaphors, gesticulate aptly and so on. Almost anything he may do, verbal or non-verbal, can on occasion prove successful. To go on to ask how these methods can possibly succeed is to begin to ask how people can ever be brought to see aesthetic (and Gestalt and other similar) properties at all. (Sibley 1965/2001: 38).

Thus, for Sibley and Wittgenstein, there is no *one* method of how we ought to do philosophy, but rather we employ a range of different tools that fit the task at hand; *whatever it takes*.

What Tanner has to say about this is found primarily in his symposium-piece 'Objectivity and Aesthetics' (1968) (which is a response to Sibley's account of aesthetic concepts as taste concepts on a par with secondary qualities) and his treatise on Nietzsche (1994). As Derek Matravers (2003) points out, Tanner's problem with art is that, in engaging with artworks that endorse alternative moral outlooks, 'we sometimes find ourselves fictionally assenting to moral properties we think are actually false' (Matravers 2003: 101). A central feature of Tanner's treatise on Nietzsche, in turn, is the recognition that moral philosophers, when presenting themselves as studying a specific issue in moral philosophy, are in fact always relying on background beliefs about the world that are, themselves, *contestable*. Here is the conclusion that Tanner (1964) draws from contestability in relation to Foot's neo-Aristotelian response to Hare:

> …whether we are confronted with a concept to which the family-resemblance treatment is appropriate, not only as regards the observable properties which the object designates […], but also as regards the function which it fulfils or the purposes it serves […], then there might arise legitimate differences as to which things instantiate the concept better than others. […] A dispute about whether one thing is better than another, which is unsettlable, when we have all the information about the things themselves, must become a dispute about the purposes which those things serve; a dispute, in other words, about ends, and thus about the nature of the beings who have those ends. And neither Mrs. Foot nor anyone else has done much more to show that her account of knives or plants works there (Tanner 1964: 197-8).

One sometimes hears the objection that Foot's adherence to Wittgenstein's descriptivist methodology

(which is also found in Aristotle's naturalism) 'leads her to a kind of critical social theory' (Hacker-Wright 2013: 150). Hacker-Wright (2013: 150) argues that, although she has substantive disagreements with Hare and other consequentialists, 'in representing her work as a morally neutral, conceptual project of uprooting "intruding" philosophical theories', Foot appears vulnerable to the criticism that moral disagreement can stem from a difference in worldview questioning the very conceptual foundations of a given moral outlook – an objection that Iris Murdoch raised against Hare in her 1956 symposium piece 'Vision and Choice in Morality'. This is precisely Tanner's worry with Foot's argument; the extent to which her arguments are taken as definitive may itself depend on whether her audience shares her substantive moral commitments as well as implicit views of other matters. As Tanner rightly notes, Hare need not deny Foot and Austin's (1961) claims about there being specialised thick moral concepts. What he rejects is the idea that such concepts may be tied to the *conception* of what the situation is – tied to it in the sense that understanding what normative purport a moral situation has might be unattainable to people who rejected or withheld certain evaluative concepts. Tanner writes:

> All that Mrs. Foot succeeds in showing is that, given the meaning of certain evaluative terms, and granted that we are prepared to employ them at all, then we are not entitled to apply or withhold them in accordance with our chosen criteria. But Hare's point is prior to this: it is about our preparedness to make certain evaluations in the first place; or about, if you like, what our moral vocabulary is to be. Mrs. Foot's argument only operates, given a moral vocabulary. But implicit in her argument, and occasionally explicit too, is the much more important and dubious claim that we have no choice as to what to account as harms and benefits. (Tanner 1964: 198).

Tanner (1994) makes the point that individuals are rarely asked to fictionally assent to *single* moral statements; rather, moral statements come in the expression of a comprehensive worldview, a vision of the actual world that shapes precisely what one takes to be salient and not in moral disagreement. Tanner's point, if I am right, echoes Iris Murdoch's (1956) view that fundamental moral disagreements may be more a matter of differences in structure of competing visions, such that one party cannot even see how the other 'goes on' to apply the term in question to new cases, or what might be the point of doing so. This, I maintain, is also the key to the conclusion Tanner draws from his discussion of Hare and Foot, that "we need to study the structure of a whole system of morality in order to understand the postulated relationships between fact and value, and basic and derived values, which are to be found in it, rather

than studying isolated statements or fragmentary arguments which might occur within several quite different total moral outlooks" (1964: 199). Crucially, worldviews are *comprehensive* outlooks on reality, an unruly mix of evaluative and non-evaluative claims in complex interaction as a whole. To illustrate, consider the difference in structure between a Nazi outlook upon the world and that of a Christian vision. As Matravers notes, while a Nazi worldview 'will include a raft of claims about genetics and history, and views about the sublimation of individual goals to the greater destiny of the state, a Christian worldview will include claims about historical events, together with views concerning the primacy of faith, hope and charity' (2003: 101).

Once we take seriously the suggestion that the morally relevant facts cannot be accessed except *through* some perspective, an alternative to the conceptual map with which we started begins to emerge. On the new model, 'objectivity' is no longer treated as an opposite, mutually exclusive, category to that of the 'subjective' and 'particular' aspect of the discerning moral judge. And the reason is that moral judgement (and the worldviews of historical individuals more generally) is no longer theoretically construed as mere *opinion* to be checked against universal moral standards but rather the very means for giving a verdict on alternative sources and balancing their relevance to the particular case at hand.

I end with some concluding remarks about the wider significance of the present picture in elucidating the use of concepts such as value and perspective more generally.

IV. CONCLUDING REMARKS

I have argued that the emphasis placed on context that is present in both Tanner's and Murdoch's accounts of value experience as always already structured by the concepts and parochial sensibilities at one's disposal effectively declares content-involving (and so rationality-involving) phenomena in human life to be inseparable from point or purpose. Nothing is valuable from 'the point of view of the universe'; value is always *value for* us (Dancy, 1993: 162). Tanner himself seems to assume that the emphasis on point or purpose must presuppose that facts about the valuer enter into the reflective explanation of the truth conditions of ethical or aesthetic claims in ways that render them radically perspectival. But this conclusion is premature: the general idea that evaluative claims are 'perspectival' is ambiguous between a number of readings that we should be careful to distinguish.

One option is to say that content and human-involving interests are interdependent: neither can be understood except in connection with the other. As Alan Thomas puts it, 'we respond to value and yet everything relevant to our subjective [human] perspective can bear on the process of evaluation and hence what those eudaimonistic values mean *for* us' (2012: 150). Thomas (2012) maintains that the correct way to conceive of this value is, indeed, presuppositionally. It does not enter into the truth conditions of an evaluative claim that such claims are relativized to the human standpoint.[7] By contrast, the picture that troubles Tanner opens the door to something more: to the prospect that we can see value content as *determined* by independently specifiable conceptual frameworks, patterns of attention, or on a larger scale, generic socio-political cultural narratives that are discernible in public discourse. In so far as the promise of a reappraisal of Tanner's account lies in such a reduction of meaning/value to a perspective, it is a new paradigm I think we should resist. And the reason is that we should distinguish conditions on the valuing subject from conditions on the associated value.

Such reorientation of focus makes available a distinctive mode of criticism, in which claims to 'objective' meaning *in* conceptual frameworks are criticised not as false *per se*, but as failing to yield the insight about the problem of objective meaning it was the point of those claims to provide. The conceptual framework of one's 'life-world' can reveal (or obfuscate) the object's meaning, it does not determine the object's meaning. To think otherwise would be to conflate what is represented with the representation.

To make good this claim we may follow the basic tactic of Adrian Moore's (1997) defence of 'absolute representations', representations that can be added without danger of conflicting points of view, and distinguish between the conditions of the production of a representation on the one hand and 'the role that the representation can play in such process as indirect integration' on the other (Moore 1997: 89). The central claim would be that the perspectivalness of the *production* of a representation, expressive of an answerable stance upon the world that (at least in the evaluative case) includes the history of whatever conceptual apparatus that is used in it, has no effect on the stance-independence of the latter.[8]

Just how we should best understand the relation of the parochial to that of an absolute conception of the world is something that I leave open for future work. The claim here is simply that the "producer" of an evaluative representation has a point of view *operative* in producing it; the context of the agent *betrays* a stance

7. Thomas (2012: 150) gives the following example: 'Postboxes are not red for humans; postboxes are red. In the latter claim the perspectivalness of colour discourse as a *whole* is presupposed' – and similarly for the notion of value relative to our human perspective.

8. Moore writes: "One attractive feature of this tactic is that it leaves considerable room for concession whenever anyone insists on the parochial, conditioned, nay, perspectival character of any act of producing a representation. They are right to insist on this, if it is properly understood. Apart from anything else, any act of producing a representation in an *act*, and agency itself is impossible without some (evaluative) point of view giving sense to the question of what to do. But one possible thing to do is to represent the world from no point of view." (Moore, 1997: 89)

upon the world. This preserves a critical stance, in as much as we are now in a position to hold that the *route* to ethical truth will be stance-dependent, shaped by one's conceptions, and yet think of competing conceptual frameworks as offering different perspectives on the object of inquiry – without thereby reducing meaning and truth to a perspective.[9]

REFERENCES

Austin, J. L. (1961), *Philosophical Papers*. Oxford: Oxford University Press.

Baker, G. (2006), *Wittgenstein's Method: Neglected Aspects*. London: Blackwell.

Beardsmore, R. M. (1969), *Moral Reasoning*. London: Routledge and Kegan Paul.

Currie, G. (1993), 'Interpretation and Objectivity', *Mind* 102: 413-428.

Dancy, J. (1993), *Moral Reasons*. Oxford: Blackwell.

_____ (2004), *Ethics Without Principles*. Oxford: Oxford University Press.

De Gaynesford, M. (2010) 'The Bishop, the Chambermaid, the Wife and the Ass: What Difference Does it Make if Something is Mine?', in J. Cottingham, P. Stratton-Lake and B. Feltham (eds.) *Partiality and Impartiality: Morality, Special Obligations and the Wider World*. Oxford: Oxford University Press.

Dretske, F. (1971), 'Conclusive Reasons', *Australasian Journal of Philosophy*, 49: 1-22. Reprinted in Dretske, F. 2000: *Perception, Knowledge and Belief: Selected Essays*. Cambridge: Cambridge University Press.

Foot, P. (1959), 'Moral Beliefs', *Proceedings of the Aristotelian Society* 59: 83-104.

_____ (1961), 'Goodness and Choice', *Aristotelian Society Supplementary Volume* 35: 45-60.

_____ (1972), 'Morality as a System of Hypothetical Imperatives', *Philosophical Review* 81: 305-316.

Geach, P. (1956), 'Good and Evil', *Analysis* 17: 33 – 42.

Hampshire, S. (1967), 'Logic and Appreciation', in W. Elton (ed.) *Aesthetics and Language*. Oxford: Blackwell. pp. 161-186.

Hare, R.M. (1972), *Essays on the Moral Con-*

ncepts. London: Macmillan.

Hacker-Wright, J. (2013), *Philippa Foot's Moral Thought*. London: Bloomsbury.

Loeb, D. (1996), Generality and Moral Justification', *Philosophy and Phenomenological Research* 56: 79–96.

Matravers, D. (2003), 'Fictional Assent and the (so-called) "Puzzle of Imaginative Resistance"', in Matthew Kieran and Dominic McIver Lopes eds. *Imagination, Philosophy and the Arts*. London: Routledge. pp. 90-105.

Mayo, B. (1958), *Ethics and the Moral Life*. New York: St Martin's Press.

McDowell, J. (1998), 'Criteria, Defeasibility, and Knowledge', in his *Meaning, Knowledge, and Reality*. Cambridge, MA: Harvard University Press.

Moore, A. W. (1997), *Points of View*. Oxford: Oxford University Press.

Murdoch, I. (1956), 'Vision and Choice in Morality', *Proceedings of the Aristotelian Society, Supplementary Volumes* 30: 32-58.

Read, R. and Hutchinson, P. (2008), 'Toward a Perspicuous Presentation of "Perspicuous Presentation"', *Philosophical Investigations* 31: 141-160.

Scriven, M. (1966), *Primary Philosophy*. New York: McGraw Hill.

Sibley, F. (1959), 'Aesthetic Concepts', *Philosophical Review* 68: 421-450. Reprinted in J. Benson, B. Redfern and J. Roxbee Cox eds., *Approach to Aesthetics: Collected Papers on Philosophical Aesthetics by Frank Sibley*. Oxford: Clarendon Press, 2001. Pp. 1-23.

_____ (1965), 'Aesthetic and Nonaesthetic', *Philosophical Review* 74: 135-159. Reprinted in Benson, Redfern & Roxbee Cox *op. cit.* pp. 33-51.

_____ and Tanner, M. (1968), 'Objectivity and Aesthetics', *Proceedings of the Aristotelian Society, Supplementary Volume* 42: 31-72.

_____ (1983), 'General Criteria and Reasons in Aesthetics', in M. C. Beardsley and J. Fisher (eds.), *Essays on Aesthetics: Perspectives on the Work of Monroe C. Beardsley* (Philadelphia: Temple University Press): 3-20. Reprinted in Benson, Redfern & Roxbee *op. cit.* : pp. 104-118.

Tanner, M. (1994), *Nietzsche*. Oxford: Oxford University Press.

Thomas, A. (2005), 'Reasonable Partiality and the Personal Point of View', *Ethical Theory and Moral Practice* 8: 25-43.

_____ (2012), 'Nietzsche and Moral Fiction-

alism', in C. Janaway and S. Robertson eds., *Nietzsche, Naturalism and Normativity*. Oxford: Oxford University Press.

Travis, C. (2005), 'A Sense of Occasion', *Philosophical Quarterly* 55: 286-314.

Williams, B. (1979), *Ethics and the Limits of Philosophy*. London: Penguin.

THE NEUTRALITY OF THE MORAL PHILOSOPHER

by Mary Midgley

MARY MIDGLEY was born in London in 1919, and studied at Oxford. In 1942 she became a civil servant and teacher, before returning to Oxford in 1947 to do graduate work. In 1949 she spent a year at Reading, before moving to Newcastle; for a few years she gave up teaching while she raised her sons, but she taught in Newcastle between 1962 and 1980, and remained as a vital force in the subject there long after the Philosophy department was closed in the mid-eighties. She published her first book *Beast and Man* in 1978 ('I wrote no books until I was a good 50, and I'm jolly glad because I didn't know what I thought before then'), and has since written widely on topics rather difficult to capture using the taxonomies of contemporary academic philosophy: human nature, ecology, how to live well, and the relationship between science and religion.

by Mary Midgley

Bones.	And murder?
George.	And murder, too, yes.
Bones.	He thinks there's nothing *wrong* with killing people?
George.	Well, put like that, of course . . . But, *philosophically,* he doesn't think it's actually, inherently, wrong in itself, no.
Bones, (amazed).	What sort of philosophy is that?
George.	Mainstream, I'd call it. Orthodox mainstream . . .
	. . .
George.	. . . In the circumstances I was lucky to get the Chair of Moral Philosophy. Only the Chair of Divinity lies further below the salt. . .

<div align="right">Tom Stoppard, Jumpers</div>

I SHALL DISCUSS THE IDEA that moral philosophers ought to be neutral as between moral views, and I shall suggest that this idea, however attractive, makes no real sense. I start from Stevenson's admirably clear statement of it on the first page of his book *Ethics & Language* (the italics are mine):

> This book deals, not with the whole of ethics, but with a narrowly specialized part of it. Its first object is to clarify the use of *the ethical terms*—such terms as "good", "just", "ought" and so on. Its second object is to characterize the general methods by which ethical judgments can be proved or supported.
>
> The purpose of an analytical or methodological study, whether of science or ethics, is always indirect. It hopes to send *others* to their tasks with clearer heads and less wasteful habits of investigation. This *does not require the analyst, as such, to participate in the enquiry that he analyses.* In ethics, any direct participation of this sort might have its *dangers.* It might deprive the analyst of his *detachment,* and *distort a relatively neutral study into a plea for some special code of morals.* So although normative questions constitute by far the most important branch of ethics, pervading all of common-sense life, and occupying most of the professional attention of legislators, editorialists, didactic novelists, clergymen *and moral philosophers,* these questions must be left

here un-answered. The present volume has the limited task of *sharpening the tools which others employ*.

That philosophers must be fair and detached is clearly right. But what form should this detachment take? Two questions arise:

> 1). Between what are we to be neutral? What is a Moral Code?

This term can cover a wide range of things, from mere differences of practice *(e.g.,* we eat our dead, you burn yours,[1] but in both cases the point is entirely to do them honour) to wide differences of conceptual scheme, which may entirely change the meaning and value of acts even when both parties approve them. *(E.g.,* crudely, you measure everything by honour, we by God's will, they by the customs of their ancestors—concepts which lead, with increasing sophistication, into moral philosophy.) The distinction can be simply made by asking: is the notion that "we must respect each other's moral codes" or "no moral code is perfect" itself part of a moral code or not? Stevenson presupposes it, but not everyone would agree. Does talk of "direct participation" as being a *danger* make sense on all moral codes?

For clarity, I shall here use the term Moral Code only in the first sense, for a set of outward practices or behaviour patterns, and shall call the range of conceptual schemes by which we explain or justify them Moralities. I shall not now raise the question of the relation between them.

> 2). Why should Detachment lead to the primacy of enquiry about a few terms used in Ethical (or Moral) Judgments?

Stevenson assumes that the predicates he mentions are *the* "ethical terms". Judgments using them are the first concern of philosophers in this field. And it later turns out that the difference between these terms themselves is negligible; *good* is the only one that matters. The "narrowly specialized part" determines the whole. *Good* supplies the one invariable form for moral discussion; that done, there are an indefinite number of possible contents *(moral codes)* which will all fit it equally well. Any analysis given of this one form is therefore neutral as between them.[2]

For making this one form crucial he needed to give no reason; this habit had been accepted in Anglo-Saxon philosophy since its introduction by Moore and Prichard. Nor did Moore and Prichard themselves give any very clear reason for it. It looked plausible in their day on certain formal grounds which, I shall suggest, are now discredited, and on moral grounds which are respectable, but quite inadequate.

1. Herodotus III. 38.

2. Compare Hare, *Freedom & Reason* p. 97; "Thus ethics, the study of the logical properties of the moral words, remains morally neutral (its conclusions neither are substantial moral judgments, nor entail them, even in conjunction with factual premises)." And p. 154: "Our moral language is neutral as between ideals".

I shall now look further at the notion of a Morality.

We all have these conceptual schemes; the question is, how far should we leave them in the cloakroom while we are doing philosophy? The ideal here, I shall suggest, is not Neutrality but Impartiality. By Impartiality I mean: readily admitting that there *are* other schemes besides the one we favour; doing our best to understand them, to state them fairly and examine them rationally. Neutrality, by contast, would be withdrawing outside the battle, saying nothing but what is compatible with both opposing arguments. To be neutral about how far (for instance) fair controversy is a good thing seems impossible, since we have to decide to argue either fairly or unfairly. Being *impartial (i.e.,* pointing out carefully the dangers and drawbacks of fairness) is not at all impossible; Mill did that frequently in the *Essay on Liberty.*

With this sense of *morality* I ask; does taking sides between moralities also fall outside Philosophy?

In theory, Stevenson's programme leaves room for it inside, since moral philosophers do appear at the end of his list of people who can properly interest themselves in question other than the analysis of the moral judgment. But in practice, both Stevenson's own analysis and the rival ones opposed to it exclude this, because that analysis is supposed to supply all the formal tools needed to handle the various contents. Contending schools of moral philosophy therefore take their names from views on that one question—Intuitionism, Emotivism, Prescriptivism.

But why should that question ever arise?

I shall argue, by contrast, that there is no such single, central form to give rival analyses of. People who claim to be analysing it in different ways are in fact bringing forward *different* forms belonging to different aspects of morality. To emphasize contemplation, feeling or action is to take up a *moral* stance, not a logical one. Each brings out the importance of one formal element, as against others which they ignore or deprecate. *You can be as biased about form as about matter.* The impression of detachment which each genuinely has springs simply from the well-known law that one's own position on such matters does not look like a particular position at all, but something obvious and universal. We have to resist this; not elevate it into a principle.[3] Various philosophers, starting with Philippa Foot,[4] have suggested breaching the monopoly by studying a much wider range of concepts, which method Hare dismisses as "not sufficiently general (in the sense in which Newtonian mechanics is not sufficiently general)".[5] But in trying to be an Einstein, one has to make sure, first that

3. Thus, for instance, since the British Middle Class tends to live largely in the future, we find it plausible to treat calculation of consequences as part of the central moral form. But it would not look so to a Samurai.

4. *E.g.,* at the end of "When is a Principle a Moral Principle?" *P.A.S. Supplementary Volume,* 1954.

5. *F. & R.,* p. 25.

one is not just being a Pre-Socratic (extrapolation of a favourite pattern really is not a virtue) and, second, that Einsteins are what one's subject needs. Moral language (and indeed language altogether) seems more likely to need Darwins. What Darwin did was to get *rid* of a perfectly general, simple view on the Origin of Species, substituting a highly complex programme which is still being worked out to-day. He was simple only in that his principles made sense with those of other branches of knowledge, that he broke the isolation of his subject.

To see how odd it is to expect moral philosophers to be antiseptically neutral, we can glance at the errors of the unregenerate dead. All serious moral (and political) philosophers have regarded their work as something which was meant to affect people's lives.[6] They all thought of formal analysis as a way of dealing with practical difficulties about choosing a way to live—a long way, certainly, and even in a sense "indirect", but certainly not vicarious (not just for "other people") and in the end the best way. And this is as true of Empiricists as of Rationalists. Even Hume meant to change people's lives where he thought it necessary, *e.g.,* in his attacks on Superstition and the Monkish Virtues;[7] where he didn't think it necessary, he meant to justify them in staying put. Mill's books were a part of his active work of reform. And Moore considered that the "ultimate and fundamental truth of moral philosophy" was his discovery that "by far the most valuable things which we can know or imagine are certain states of consciousness, which may be roughly described as the pleasures of human intercourse and the enjoyment of beautiful objects".[8] Not much neutrality there.

The interesting question is: were they all just making a crude mistake? Did they simply fail to notice that they had started their enquiry the wrong way, and were therefore, like old-fashioned tradesmen, carrying on two separate businesses behind the same board? (Joiner & Undertaker; Haberdasher & Florist; Moralist & Moral Philosopher . . . ?) Do modern techniques make it possible for us simply to lift their analytic contribution from their recommendations, and use it, not (as they supposed,) to reach a particular morality (namely, the best one), but to understand equally *all* moralities, to an indefinite number? Or should we be more drastic still, and think of modern analysis as an improved technique putting them out of date altogether?

Again, are the non-professionals listed by Stevenson really working outside the subject? Can, say, Tolstoy, Dostoevsky, Swift, Proust, Mann tell us nothing about what it might *mean* to find something good?

I think the whole subject important, not only within

6. See *e.g.,* Plato, *Republic* 352d; Aristotle, *E.N.* 1.3, and especially Kant, *Groundwork,* end of Chapter 1. (Paton's tr. pp. 69-70)

7. *Enquiry* §159 & §219.

8. *Principia Ethica* ch. 6, pp. 188-9.

ethics, but in relation to the more general one: *what is a subject?* What individuates any branch of learning, and particularly a branch of philosophy? Need there be a branch called Moral Philosophy? if so, why? is that the best name for it? how far does it stretch and what are its neighbours? I shall answer:

(a) that it is indeed right to look for a distinguishing mark in the *form* of arguments involved, not just in a difference of objects, but that,
(b) the form will not necessarily be a simple one, and,
(c) in the case of morality it certainly is not.

To expand:

(a) The important differences of subject are indeed those of form not matter.[9] Thus, Physics is not the study of a particular set of things called Physical Objects, biology of different Objects and Psychology of different ones again; they are all studies of the one world, but from different points of view, using different concepts to select and concentrate on different patterns in it. None eliminates the others; none is "fundamental" in the sense that its objects are *the real* ones. They can be mutually related, but not reduced. Thus again, the same poem can be studied by the Critic, the Historian, the Linguist and the Psychologist; their difference will be in the points that concern them; in their concepts and the general framework into which they fit the poem. If moral philosophy needs to exist as a subject, it should indeed serve such an interest and have such a characteristic framework.

But (b) *what is odd about recent ethics is the assumption that this form must be the simplest one available,* namely the "Moral Judgment." I go back here to Stevenson's parallel with the Philosophy of Science. Is this best approached by finding a single simple form common to all propositions of all the sciences (say, just "*x* is a fact"—or might we include "*y* is a sound hypothesis" as well?) and remaining neutral on all questions of the relations and interpretations of facts? Should a philosopher of science always forswear tracing more detailed forms, on the ground that those are valid only in particular fields, or favouring one particular hypothesis (say, that of Copernicus or Darwin) with its associated concepts, rather than another? If he does that, how will he "send others to their tasks with clearer heads and less wasteful habits of investigation"? Locke's plan of acting as an "underlabourer" in the cause of science[10] *involves* direct participation; an underlabourer who leans on a post outside the site telling you what it all amounts to is a doubtful asset. Locke wanted him to cart away the rubbish. But to do that he has to take sides against it. *Impartiality* in this field means examining fully all sensi-

9. There is of course a *relation* between form and matter, in that certain forms are evolved to handle certain aspects of the world, and cannot be used on others. You cannot use biology to talk about stones, nor physics to talk about human history. It follows that I am in agreement with those (P. Foot, Warnock *etc.*) who want to say that morality has a distinctive subject-matter. I differ from them (if at all) only in emphasizing that that involves complex distinctive forms as well which are more directly the concern of philosophers. *Even* on Form (supposedly its strong ground) the current view is indefensible, (See G. Warnock, *Contemporary Moral Philosophy* pp. 52 ff.)

10. *Essay on Human Understanding,* Epistle to the Reader.

ble alternatives, giving the reasons why one hypothesis is better than another, and saying for what purposes it is so. Ptolemaic epicycles are apparently still used for some purposes of rocket-launching, and Einstein does not for all purposes displace Newton. In order to understand "*y* is a sound hypothesis", you have not only got to have heard of such procedures, you have got to *accept or reject* the principles on which they work. If you are, as Stevenson says, to "sharpen the tools which others employ", all parties must be agreed on just what *kind* of sharpness is called for—razors, hatchets, saws or spades? and must know better than to sharpen the hammer and the foot-rule. For this, it is commonly an advantage if the sharpener *does* sometimes use the tools himself.

Another example would be the Philosophy of History. We might rule here that only one central thesis mattered, namely, that the unit of history was the simple categorical proposition in the past tense, stating a particular historical fact. All views on the relations of these atomic facts, all general concepts providing alternative ways of grouping or selecting them, and particularly all modal concepts like historical necessity, would be dismissed as merely points of content, without philosophic interest . . . ?

Or in the Philosophy of Religion, we might say, "Our concern here as philosophers is simply to define *the religious terms*, namely, 'God' and 'Belief. The formal work is then done; there remains only the non-philosophic, purely material question, *which* god to believe in . . . ?"

The first two suggestions are not actually a fantasy; when we attack the prejudices and preconceptions of particular historians or scientists we may have something like this pattern in mind. But nobody would seriously aim at completing the programme. We shoot down *false* preconceptions; without preconceptions of any kind, thought would stop. If all facts were equal and all groupings tabu, there could be no selection and nothing could be said at all.

My parallels may seem far out. All the same, this is just the way in which, from *Principia Ethica* onwards, people have taken it that the important thing about morality was its having a standard formal unit, the Moral Judgment. Looking back, we can see what they meant by it. Moore and Prichard saw this Judgment as a simple SP proposition using a single characteristic (Good for Moore, Right for Prichard); for this predication both held that "no relevant evidence whatever can be adduced".[11] Both thought that their predicate, the *fundamental notion* of morals, must be something intelligible quite on its own, not just a notion receiving special

11. *Principia Ethica* Preface, p. viii.

emphasis. Each could see the faults of the other's horse; neither realised that the whole race was impossible. *But clearly no concept is isolated in this way.* One could as well argue about which is the fundamental concept of History—pastness, sequence, or human affairs? Or again, is symmetry found in the right or left half of the pattern? does fertility lie in the seed, the plant, or the species? I am sorry to sound like an Idealist, but there actually is such a thing as vicious abstraction, and this is it. *Concepts work in sets;* you can only use one alone by quietly presupposing the others. To talk of *right* is to talk in a context of order, about responsible beings with definite claims and duties, and probably in some sense with free will. To talk of *good* is to talk in a context of purpose, about beings with a definite set of tastes, wants and capacities. Another step or two, and the contexts will meet (we ask: is this the right purpose? is that a good order?). These frameworks are formally quite different from the simple context of colour-words. And even colours have the spectrum.

The Intuitionists of course did not pretend to Neutrality. But from their isolation of the Moral Judgment, given changes in the world, the Neutrality claim follows.

What is the alternative to picking out such a single form? Clearly, it is to investigate the various forms of moral thinking which are actually in serious use at any time, particularly those likely to give trouble. Thus, if we had to make a list of "the ethical concepts" to-day, it would be right for a start to include terms like importance, need, work, fulfilment, meaning, normality and sickness, maturity, happiness, society, freedom, intelligence, alienation, despair, exploitation, commitment, humanity. Of course some of these notions might turn out to be bogus, literary or tendentious, but then it is the business of philosophy to show this, not to presuppose it. And what it means to be bogus, literary or tendentious is just the kind of question which needs philosophic treatment; these notions are not covered by "nonsense". Of course, too, it would mean that there could be no short, exhaustive list of such concepts. Other subjects have no such lists, only a group which are prominent and typical, and which change from time to time with the state of the argument. It would also mean being willing to relate moral arguments to those arising in neighbouring fields, not just in other branches of philosophy, but in, for instance, psychology, anthropology and history. There is not a hard line dividing moral thinking from a homogeneous set of things called *facts,* because, for a start, there is no such homogeneous set. *Brute Facts*[12] have to be organised progressively (in every field, and not just in that of morals) under increasingly intelligible headings before we can decide what they are to count as. By the

12. See Elizabeth Anscombe on *Brute Facts,* Analysis 1958, and in "Modern Moral Philosophy", *Philosophy,* Jan. 1958. Also an excellent discussion in Kovesi, *Moral Notions,* ch. I.

time they come up for moral judgment, the job is largely done.[13] The right question is always Hume's question,[14] "how did you get *those* values from *these* facts?" not "how did you get values from facts at all?" There are plenty of bad ways. *E.g.,* very roughly and provocatively; Evolutionary Moralists have got hold of thoroughly unsuitable facts to support their value conclusions; so have people who argue *just* from custom, authority or fashion. Hume's talk of Desire, and Bentham's of Pleasure & Pain, use much more suitable facts, but use them very crudely. People like Aristotle, Butler and Mill who go beyond such notions to give a fairly convincing account of human nature, showing the *sense* of our desiring what we do, are arguing rightly, though each argument is no doubt still over-simple. In ordinary life we all make this distinction; slapping the title *naturalist* right across the range is just careless. Some arguments (particularly very simple reductive ones) from premises lying more or less outside morality to moral conclusions are invalid, as Moore showed. Others, as many recent philosophers have pointed out, are sound.[15] It is philosophy's business to distinguish these kinds. And the frontier of morality itself varies with the kind of morality you accept.

Let us look more closely at the original reasons for isolating the Moral Judgment from supporting evidence.

1. Formal Reasons. The Empiricist Programme of the day was to find a set of standard units into which all meaningful discourse could be resolved. Moore and Prichard gave Moral Judgments the form of isolated intuitions *(x* is right, *x* is good) strictly parallel to that paragon of tidiness the sense datum proposition *(x* is pink). That fitted them for their place in the analytic scheme, and did so without the obvious distortions of Reduction.

With this plan in mind, they naturally thought that other people were darkening counsel if they suggested other analyses which, though still formal, did not fit this mould. That went for Kant,[16] whose scheme was formal enough, but involved a whole network of other concepts (choice, will, law, responsibility, freedom *etc.)* surrounding the central one of a Rational Being. He was untidy too in refusing to dissolve the moral imperative into any sort of indicative proposition. There was similar trouble over Aristotle,[17] who defined Good by reference to an Organism, its nature, aims and desires, and to the kind of reality which its various aspects may be said to have. This is still formal, but again it is complex. The SP judgment rests on the facts about the organism in question, and, when made, it has, as Moore saw, a different force BECAUSE OF ITS PLACE IN THE FRAMEWORK. In neither system can moral judgments be fundamental or atomic.

13. Thus, Hare says that the Liberal bases his case on *"the facts"* but the Fanatic on *"careful selection of the facts"*. So how did the Liberal get his? Wholesale? Our structural concepts determine what we count as *the facts*—which is why they are the right place to start our enquiry. *(F. & R.* p. 181.)

14. *Treatise,* Book III, part I, end of section I.

15. *E.g.,* P. Foot, Moral Arguments, *Mind* 1958: G. Warnock, *Contemporary Moral Philosophy* pp. 69-70, E. Anscombe and Kovesi, *loc. cit.* Also Searle, "How to Derive Ought from Is", *Philosophical Review,* Vol. 73 (1964).

16. *Principia Ethica,* ch. IV.

17. Prichard, "The Meaning of *Agathon* in the Ethics of Aristotle", in *Moral Obligation.* Moore, "The Nature of Moral Philosophy," in *Philosophical Studies.*

Someone, then, who in advance sets certain standards of simplicity may simply reject these analyses on ground of extravagance. This Moore did. The Logical Positivists then slung Moore's non-natural facts out after them on the ground that in a proper Empiricist universe these too failed to earn their keep. But for this move to work, the slinger ought to have a satisfactory simple unit of form to which all meaningful discourse (including morals) *can* be reduced, making all the complex forms dispensable. (It is no good complaining of extravagance if the goods bought turn out to be necessary, and cannot be bought more cheaply.) As it turned out, plenty of other forms besides moral ones failed to fit the Procrustean bed. So that to-day most people have dropped the search for standard atoms of discourse. They do not try to delimit the Meaningful sharply. Instead they study and compare the various ways in which words have meaning. We no longer have to improvise a narrow standing-ground for Ethics in a framework devised exclusively for Science, to invent "the place of Value in a world of Facts". In this situation, it is surely right to wonder, as Geoffrey Warnock does.

> How could it possibly have been supposed that moral discourse, in all its almost endless diversity of form and contents, must consist essentially and always in the performance of any *single* speech-act?[18]

And if it had done so, should it not count as a minor twig on the speech-act branch of philosophy? Why *courses* in it? Why *professors* of it?

2. Moral Reasons. Some people believe that complex analyses of morality lead to wickedness, or at least to a waste and confusion that hinders goodness. Both Moore and Prichard certainly thought in this way. I have no space here to discuss Moore's position, which was much more subtle and less negative. But it was Prichard's example that was followed.

It was Prichard who first convinced philosophers that *not* doing moral philosophy was the right way to do it. It was also he that remodelled the style of empiricist ethics, exchanging the powerful, subtle instrument of Butler and Hume for something more like the rattling of dried peas in a can, and thus avoiding the embarrassment of having any non-academic readers. Prichard believed that *all* positive moral argument, all endeavour to explain the moral code by an underlying morality, was self-deceiving evasion, a web woven to blind us to the surd and brutal fact of duty.[19] Now there are some cases where this monstrous exaggeration is in place as a corrective, namely, where good clear principles are actually present, but the will is weak. The same chance factor which later made Sartre so influential helped Prichard here—he was

18. Warnock, *C.M.P.* p. 34, and 39-42.

19. See especially the very influential article "Does Moral Philosophy Rest on a Mistake?", first published in *Mind* for 1912.

talking chiefly to intellectuals, who often do have this trouble, and are easily made to feel guilty about it. But where principles are in fact bad or confused, strong will is not what is needed. Mankind in general plainly thinks too little rather than too much, and does not need philosophers to stop it thinking.

His position has also a less trivial ground in the drastic changes in the world which were already under way at the beginning of the century. Prichard hoped to save morality unharmed by unscrewing it from the facts, and treating the changes as merely factual changes. But this cannot work. *All gross changes in the facts demand new concepts,* raising new problems which simply cannot be handled through the old set. Because dogmatic conventionalism cannot meet these demands, it ends in moral bankruptcy.

Prichard's appearance of stern rigour, therefore, was of the kind you get by throwing all the real problems into the cupboard and locking the door. If you assume his moral position (a hatred of indecision and an acceptance of conventionality), you do get a reason for isolating the form of Moral Judgment. Without that, it dissolves.

The trouble then with Moore and Prichard is not that they took sides, that they recommended their own values, but that they took that recommendation for granted by failing to deal properly with rival and conflicting values. Their successors did not supply this need. Instead of bringing Moore's anti-naturalism up to confront in detail the live naturalistic and metaphysical views which were its natural opponents, instead of developing it to deal properly with Marx, Freud, religious thought, and the various kinds of modified Utilitarianism we all largely live by, they simply took it as settled, ignored the outer scene, and tried to make it more thorough. Moore's mistake, they decided, lay in not making his minimal particle minimal enough. Emotivists and Attitude Theorists looked for a simpler form still, common to all moral judgments. But since this simpler form is common to all sorts of other judgments which no-one would call moral, it gets hard to say what is special about moral ones, or why the subject should exist. (Stevenson's suggestion that *moral* approval is marked off from liking by "a rich feeling of security"[20] is no great help. Stevenson indeed might well accept that the subject should really be merged in a more general one called the Philosophy of Persuasion, formerly Rhetoric.) As so often happens, the HCF is insignificant. He who, on asking what the preacher said about Sin, is told, "He was against it", doesn't go away much the wiser. In what *way* was he against it, for what reasons, to what extent? how did he define it and relate it to the rest of life? what did he think

20. *Facts and Values,* p. 25.

were its *genera* and *species?* Emotivism and Attitude Theory strained all such points out of their surgically pure Moral Judgments, to make them fit any content equally well or badly. Thus they converted Prichard's machine for Standing Still Morally into one for unlimited drifting—a change which is not really surprising seeing that the machine worked by making the choice of code arbitrary in the first place. (A tortoise, withdrawn into his shell, feels secure, but in fact he can be rolled as far as anybody cares to kick him.)

The Universal Shoe had at this point become so loose as to fit nobody. Yet it is unluckily possible to be trivial and partial at the same time, and there were still, in fact plenty of places where it pinched, essential aspects of morality which it managed to exclude.

E.g.—It is often of great moral importance, not just whether we are for or against something, but *on what grounds* we are so. If I oppose *(e.g.)* aversion therapy or cruel experiments on animals, it makes a crucial difference whether I do it simply from a traumatic horror of labs., a grudge against scientists, loyalty to a movement, or a clearly thought-out view of the principles involved. Stevenson's theory is specially designed to make this difference unstatable, nor is Nowell-Smith's much better. They are like maps which deliberately leave out certain sections of road. This is the partiality of silence.

The looseness and partiality of Emotivism did in fact dissatisfy people. For that reason, in Prescriptivism the very sensible step was taken of letting Form become more complex again; moral judgments were limited to those universalizable and directing conduct. But if we go so far from the minimum, what stops us going further? *A set of universal prescriptive rules is nothing like enough to make a Morality.*

Let us look again at the parallel with Science.

How would it be if we allowed, as *the* form of scientific judgment, just universalizable assertion, and as matter just brute facts?

In mentioning this (quite close) parallel before, I charitably allowed scientists to use the form *y is a sound hypothesis* as well as *x is a particular fact*. On the present interpretation, all *y*s will be universal propositions, to form which many *x*s are assembled, and from which others are deduced. So what is missing? Only the vast apparatus of concepts necessary to make selection possible—in fact, the whole field of scientific method.

Take just one major branch of science: biology. For each part of it, we need to understand what is meant by—life, organism, organ, plant, animal, species, func-

tion, adaptation, selection, ecology, habitat, nutrition, reproduction, respiration, cell *etc.* Every particular branch of biology adds its own further set. Physics and Chemistry, Geology and Climatology have their own; biology needs many of them too. There are also the more general concepts which concern every science, such as nature, experiment, structure, and plenty of more general philosophic concepts too, such as relevance, evidence, proof. And so on.

Not until all this lot are more or less in place can a particular fact X *become part of a science.* Before that we can neither select it nor extract a meaning. "There are no plants in this box" is on its own a fact without value or consequences. (Though it should be noted that it is a long way from being a Brute Fact, which would be something more like "earth in wood here" or "dark brown all over".) To be used, it has to take a form like "no *seeds* have *grown* in this *soil,* even with *optimum conditions",* or "this *hybrid pea-plant has produced* no *offspring".* The italicized words mark the crucial part played by the structural concepts. (Note how easily this *fact* might have read, "no seeds have grown without blood-sacrifice" or "when my neighbour has looked at them" or "while I have this wife". And of course these are examples of *bad* concepts, not of no concepts at all.) What the philosophy of science does is just to relate and clarify these concepts.

Can we say that all this work concerns *content,* and that the only strictly *formal* elements involved were deduction and assertion?[21]

There could be two reasons for saying this. First, one might have a formal theory to protect which demanded this ruling. This would not be a very impressive position. Second, and better, one might think that faults connected with deduction and assertion were particularly *dangerous.* Now it is certainly true that scientists who lie, haver or commit deductive fallacies are disastrous. Faced with cases of these paralysing vices, we may hastily say that honesty, firmness and deductive consistency are all that matter in a scientist. So what do we then say of a man who has all these virtues, and lacks only curiosity, discrimination, common-sense, energy, humility and imagination? It sticks out a mile that we have taken the structure of concepts for granted.

And this is pretty much what has happened in moral philosophy. Thus, Hare's examples of possible moral principles commonly cite patterns of outward behaviour, on the model of his early example, "one ought never to smoke in compartments in which there is a NO SMOKING notice" *(LM* 176). But this has no moral bearing at all unless we understand some real moral con-

21. This seems parallel to Hare's position in *e.g., F. & R.* p. 200, that "the concepts of the naturalist are not formal".

cept (it might be consideration, obedience or fairness) under whose heading we can bring it. Before universalizing you would have to know *which* concept you had in mind, because they would give totally different sets of parallel cases. On Hare's view, King David's proper and sufficient answer to the Prophet Nathan would have been to retort, "Nonsense, I have never stolen any sheep" (2 *Samuel,* 12.7). *It is the conceptual structure that determines the morality.*

I return now to this difficult and useful term (morality) which I have used provisionally to cover everything in the way of explanatory systems and principles lying between a mere list of commanded and forbidden acts and a mere general form of moral judgment.

A morality cannot just be a list of acts that one is for or against. It is a set, and a coherent set, of reasons for being for or against them; it is a way of thinking. (1) The acts a person is for or against vary greatly with circumstances while his principles remain the same, and (2) people who favour the same action, and act alike, but on quite different grounds, differ morally.

Thus (1) "honour thy father and thy mother" will suggest totally different acts in different cultures with different ways of showing honour, or with changes like the illness or eccentricity of a parent. But the people invoking the principle can understand each other morally once they know about the difference in circumstances.

(2) If we are asked about the reason for this commandment, we may back it simply by tradition, or prudence ("what will happen to *us* if the rules crack?") or with arguments drawn from the general happiness, or with talk of justice, rights, obligations and respect for persons, or with direct consideration of the feelings involved. The outward actions could be just the same, but their meaning for those concerned, the spirit informing them and the part they play in our lives, would be quite different. That is what a moral difference is. It will not do to say (as Sartre does) that love is simply the deeds of love;[22] what use are the deeds of love if they spring from a single-minded desire for a legacy? No more is morality the acts of morality—nor, therefore, a set of precepts prescribing these. It is far more a characteristic *way* of thinking, feeling and acting. Each morality has therefore its own quite complex set of forms. Being impartial involves watching out for these different forms and relating them, not slurring over the variety. This comes out in the fact that not all forms of moral reasoning are equally suitable for explaining all moral views. If, for instance, you are trying to say *what is wrong* with subliminal advertising, slapdash prescribing of sedatives, commercial pornography or aversion therapy, you are

22. *Existentialism and Humanism* p. 41 (Mairet's trans.)

helpless if you can only say, with Stevenson, "I disapprove; do so as well", or, with Hare, "never do things like that". (Like it in what *way?*) You will need, even if you have never heard of Kant, to draw on more complex forms of argument, such as those involving distinctions between people and things, between being manipulated and choosing freely, between conditioning and teaching, desire and respect. If, on the other hand, you want to ask what is the point of some hive of ceaseless industry, like the Transport Business, or the Stock Exchange, or a factory for making plastic gnomes, you may do better, first with Mill, then with Chapter 6 *of Principia Ethica* ("whose life is actually the richer for it? and in what way?") And if you were pleading the cause of a certain sort of conscientious objector, there might even be a place for Prichard.

It is not obvious, in any of these cases, that your argument will be best understood and analysed by a "detached" philosopher who inclines to neither side. Your form and matter fit each other, just as literary form and matter do; the point of arguing in that way *is* also the point that makes the case worth presenting; if he misses the one he will be puzzled by the other. What the form does is to connect your particular disapproval with a whole general way of viewing and treating people; you are not just proscribing certain acts, but saying what is wrong with them.

Conclusion. I have tried to suggest that it is a mistake to make the analysis of *moral judgment* the central question of moral philosophy, and to suppose that it can be treated in abstraction from all moral preference. There need not be, and does not seem to be, any single, simple, logical form or speech-act characteristic of morality. There is rather a wide range of such forms, linked variously with each other and with other subjects. To concentrate attention on some forms rather than others by calling them more "fundamental" or "central" is to take sides morally, not to make a logical discovery. If we emphasize *e.g.,* feeling, obligation, command, consequence, motive or outward act, we are offering a particular morality, a certain way of viewing and treating life. We ought therefore to argue for it openly, not take it for granted, and to accept that, if people believe us, it can affect their lives. I conclude then that moral philosophers are and cannot be neutral.

There remain problems about what they are and can be. What *does* distinguish the subject from its surroundings and give it point, if not a basic formal unit? Are moral philosophers necessary?

I suggest that their job is to examine the forms of morality in general, that is, of the various and overlapping

moralities by which people live, and particularly those which are actually making trouble. I do not think we can dogmatize about how much should be included. There is an immense overlap, often taken for granted, between the lives and interests of all human beings, which means that some notions are obviously good candidates for being considered central, and some forms are found underlying a wide range of topics. But both centrality and pervasiveness are matters of degree. The moral point of view is, I suggest, the one where we ask, "What are the most important, the central things in life?" and some forms of thought do emerge as perennially suitable for these enquiries. But they are a complex set, and may always shift their roles with changes in the world. (For instance, an oncoming Ice Age would greatly shift people's moral priorities, without necessarily altering their deepest principles.) Words like *moral* are therefore inclined to behave elusively. What people choose as *the* fundamental moral form varies largely with the opposition, with what they want to deny at the time. And *moral,* like other hard-pressed words (say *art or freedom),* collects a very wide range of opposites. There is nothing to be gained by freezing it in a fossil sense.[23]

23. I have tried to trace a few of its windings in an article called "Is Moral a Dirty Word?" *in Philosophy* for July 1972. I have also stuck my neck out by outlining an account of how we *can* argue from the facts about Human Nature in an article on the Concept of Beastliness in *Philosophy,* April 1972, and continued the process in another on Games, forthcoming in *Philosophy* shortly.

IN DEFENCE OF (A SORT OF) NEUTRALITY

Commentary by Roger Crisp

ROGER CRISP is Uehiro Fellow and Tutor in Philosophy at St Anne's College, Oxford, and Professor in Moral Philosophy at the University of Oxford. He is the author of *Mill on Utilitarianism*, *Reasons and the Good*, and *The Cosmos of Duty*. He currently holds a British Academy Thank-offering to Britain Fellowship, and is working on the relation of self-interest and morality in British moral philosophy after Hobbes.

IN DEFENCE OF (A SORT OF) NEUTRALITY

Commentary by Roger Crisp

There is much to be learned from this paper, much to be inspired by, much to agree, and to disagree, with.

Midgley's main target is the view that moral philosophy should concern itself only with *neutral* (that is, evaluatively and normatively neutral) conceptual analysis. She herself prefers *impartiality*, a position of non-neutrality but readiness to assess, fairly, alternative views, along with a broader focus in ethics than she finds characteristic of post-war analytic philosophy.

Midgley begins (220-1) with a substantial quotation from the first page of the emotivist C.L. Stevenson's 1944 *Ethics and Language*. (Interestingly, as one might perhaps have expected in the free-and-easy 70s, she doesn't bother to give any publication details. She also italicizes quite a lot of the quoted passage, but doesn't mention she is doing so.) Stevenson's book had been published thirty years previously, and the emotivist 'programme', if one can call it that, was slowing down by the time of this article, awaiting its revival in the work of Simon Blackburn and Allan Gibbard. Stevenson was soon to retire from Michigan, and R.M. Hare was working on his last major book. Meanwhile, at least in the UK, significantly more interest was being shown in the naturalistic virtue ethics of Philippa Foot and the non-naturalist version of John McDowell and David Wiggins, and in those papers by Bernard Williams sceptical of the whole idea of impartial ethical theory. This may explain, at least in part, why Midgley's paper seems to have received rather little attention. Many of those to whom she was preaching had already been converted, or didn't need conversion at all.

In the quotation, Stevenson says that his book deals only with that part of ethics concerning the use of 'the ethical terms' and the methods by which ethical judgements can be supported. He goes on to claim that this kind of analysis doesn't require the analyst to participate in (what we might call) 'first-order' ethical enquiry, and that such participation is anyway to be avoided, since it may distort the neutrality of the analysis itself.

On the face of it, there doesn't seem too much to object to here. Imagine, for example, a staunch utilitarian who is trying to elucidate the sense of the word 'right'.

It may be tempting to offer a definition of the word as something like 'utility-maximizing'. Indeed that is exactly what Bentham and G.E. Moore did. But Midgley has an immediate objection to Stevenson's position (221): it is self-contradictory. His suggestion that we should be neutral is itself a non-neutral statement. (Compare what perfectionists were to say shortly afterwards about liberal 'neutralists' such as Ronald Dworkin.) This objection, however, seems off-target, if we allow that Stevenson's recommendation of neutrality is not moral but epistemic. In general, throughout the essay Midgley fails to recognize that some methodological decisions – on which concepts to employ, for example – may be epistemic rather than moral.

Nevertheless, I think the general point lying behind Midgley's position is extremely important, and has been woefully under-recognized in the history of philosophy. As she notes (221-2), Stevenson leaves unanswered the question why we should start our ethical analysis with a few terms (primarily 'right' and 'good') found in moral judgements in particular. In general, most philosophers have taken over the questions of their predecessors (consider, for example, the influence of Socrates on the ethics of both Plato and Aristotle). But they have often not bothered to provide any argument for the primacy of those questions, or the concepts they involve. Nor – as Midgley points out (e.g. 222-3) – have they noticed that the choice of these questions or concepts may not be a purely epistemic matter; it may indeed be, as Midgley puts it, 'moral', or, to use more general terms, 'evaluative' or 'normative'. (For that reason, anyone interested in the nature or tenability of the distinction between metaethics and normative ethics would be well advised to take this paper for a spin.)

Midgley claims that Stevenson didn't have to give a reason for choosing the ethical terms he did, since the approach he was taking had been accepted since it was introduced by Moore and H.A. Prichard. This is striking evidence of Henry Sidgwick's near disappearance from philosophical view after the first decade of the twentieth century, though of course by this point he had been noticed by Rawls, and indeed in 1974 a special issue of *The Monist* was published to mark the centenary of his masterpiece *The Methods of Ethics*. Prichard's focus on rightness and goodness may be a result of Moore's; but Moore's is undoubtedly a result of Sidgwick's.

How devastating an objection is Midgley's to Stevenson's project? I would say, 'Not very', though she has, as I've said, brought out something important about neutrality. What might Stevenson have said had Midgley pressed him for some justification for starting with

goodness and rightness? One response could have been: 'These are clearly central ethical concepts. Nearly all of us are interested in which things are good and which actions are right. And, though in the passage quoted by Midgley, I do also mention justice, and leave my list of concepts open-ended (something she fails to note), I am inclined to see these questions about rightness and goodness as of primary significance'.

Now this kind of response by Stevenson would not persuade everyone – and certainly not Midgley. Nevertheless, many of his readers would probably have felt roughly the same as him, and so his book might then have been valuable to them. Further, even if he were to recognize Midgley's important key point (as he should, and perhaps would, have done), there is nothing to prevent his insisting on neutrality once the initial choice of questions and concepts has been made. It may not be a neutral decision to focus on 'right', but as we see in the cases of Bentham and Moore it's a real mistake not to be neutral in analysing it. And since Midgley herself recommends non-neutrality, she herself runs into exactly the danger which Stevenson is warning against.

And, indeed, she fails to avoid it, as egregiously as Bentham and Moore. Remember that Midgley advocates not neutrality but 'impartiality'. This could be described, however, as a kind of *partiality*: one starts with a particular conceptual scheme, and then seeks to referee battles between that scheme and alternatives as fairly as possible. But imagine how Joe Frazier would have felt had he turned up for his fight against Muhammad Ali in New York in March 1971, just before the publication of this paper, and been told that the match was to be refereed 'impartially' by Angelo Dundee, Ali's trainer. Stevenson's neutrality principle is surely preferable.

Midgley herself recommends a much broader approach in moral philosophy than that of Sidgwick and his followers (223-34), described memorably by Jonathan Glover one year later in this journal as 'the preference for ethical systems in the style of the Bauhaus rather than the Baroque' (Glover 1975: 183). Midgley suggests that those who offer narrower accounts are merely analysing parts of a much larger conceptual whole (222-3). She criticizes Hare for using Einstein as a role model (that is, someone who sought a more general theory than his predecessors), suggesting that 'moral language … seems more likely to need Darwins'. Apparently, what Darwin did was to replace a general, simple view with a 'highly complex programme which is still being worked out today' (223; see 230-1). That seems rather questionable. What Darwin did was roughly what Einstein did: replace one general, simple view with a much better gen-

eral, simple view. And if that is not the way to proceed in philosophy, we need to be told why. Midgley's approach has less in common with Darwin, then, than with that of other moral philosophers. It was anticipated, for example, by Elizabeth Anscombe in her 1958 'Modern Moral Philosophy', and itself anticipates later work by Alasdair MacIntyre and Williams. All three of these philosophers favoured the paying of greater attention to the so-called 'thick concepts', in particular to specific virtues and vices, rather than developing abstract theory in the mode of Sidgwick.

I am inclined to agree with Midgley that the focus on 'moral judgement' in early- and mid-twentieth century ethics – exemplified in Stevenson, but pretty ubiquitous – was too narrow. The key question in ethics is not 'which actions are right?', or 'which things are good?', but the more fundamental Socratic question, 'how should one live?'. And, as Williams pointed out, what Socrates was seeking were *reasons* for living in one way rather than another. Moral philosophers, then, should aim at answering this question above all others. Now it may be that the best way to do this will be to spread one's net as widely as possible, offering conceptual analyses of terms such as 'importance, need, work, fulfilment, meaning, normality and sickness, maturity, happiness, society, freedom, intelligence, alienation, despair, exploitation, commitment, humanity' (226), and it may be that the aim of moral philosophers should be to 'examine the forms of morality in general, that is, of the various and overlapping moralities by which people live' (233-4). But I myself doubt it, and believe that the default choice should be Bauhaus rather than Baroque. (Nor, incidentally, do I see any problem with focusing on 'judgement', as opposed to say, feeling (232-3). Moral lives of course involve more than judgements; but moral theories are to help us decide what to believe, not primarily how to feel.)

Socrates's question is, in one way, too broad, given the essentially practical aims of philosophical ethics. What most of us want from moral philosophy is an answer to the question of how we should *act*, or, yet more narrowly, how we should *choose* or *will*. Doubtless answering that question will require us to introduce other concepts as we go; for example, to return to Midgley's list, it is hard to imagine a comprehensive ethical view that says nothing about happiness, if by that term we mean 'what is good for someone' or 'well-being'. And it is true that concepts are not 'isolated' (226). Each must be elucidated with reference to other concepts. But there are at least two good reasons for extreme caution in introducing a new concept. First, it may be unnecessary, and hence epistemically dangerous, wasting time

and creating potential confusion. Second, it will have its own cultural, evolutionary, and political history, and we should know from experience that such histories often create an illusion of universality or plausibility which can distort our thinking. The standard, but nevertheless excellent, example here is chastity. And I would also mention the remarkable case of supererogation, a notion unknown in the ancient world which became a key part of our common-sense morality through interpretations of a few passages in the Gospel of Matthew by some of the desert fathers (see Crisp 2013: 17-19).

Midgley sees moral philosophy as like a party to which all are invited, with no one on the door and no bouncer in attendance. Disagreements are much more likely to arise at such a party, and the voices of the most important people at the party will be drowned out. It may even turn into a complete *fracas*, with people shouting incomprehensibly at one another. (This is roughly the kind of state the first few pages of *After Virtue* suggest MacIntyre believes contemporary ethics to be in.) I think there is much more to be said for a Sidgwickian form of methodological parsimony. First, get clear on your most important question, or questions, and on the concepts they involve. Second, examine various answers to those questions by long, hard reflection on them, to see whether those answers are themselves clear and self-evident – that is, whether you are justified in believing them without reliance on inference. Further, this reflection should be as neutral as possible, to avoid contamination of the whole process.

Sidgwick's party will be more staid than Midgley's. It will involve a few, carefully chosen guests to start with, and later comers will be invited in only after careful vetting. All guests are under constant scrutiny and liable to be excluded at any point if they are not making a contribution. Myself, I'd much prefer to be at Sidgwick's party. But, though a great admirer of him, I have to admit that even Sidgwick was not as neutral as he had perhaps hoped to be. Quite frequently, for example, in his long and searching discussion of common-sense moral principles he will raise an objection on utilitarian grounds. But we do not have to give up on the kind of neutrality recommended by Stevenson. Rather, we should see it as an ideal, and the frequent failure of moral philosophers to achieve it should only spur us on to try harder ourselves.

REFERENCES

Anscombe, E. (1958), 'Modern Moral Philosophy', *Philosophy* 33: 1-19.

Crisp, R. (2013), 'Supererogation and Virtue', *Oxford Studies in Normative Ethics* 3: 13-34.

Glover, J. (1975), 'It Makes No Difference Whether or Not I Do It', *Proceedings of the Aristotelian Society Supplementary Volume* 49: 171-90.

Stevenson, C.L. (1944), *Ethics and Language*. New Haven: Yale University Press.

CONSTRUCTIVISMS IN ETHICS

by Onora O'Neill

Onora O'Neill, Baroness O'Neill of Bengarve, studied at Oxford and at Harvard, where she wrote her doctoral thesis under the supervision of John Rawls. She taught at Barnard College before moving in 1977 to the University of Essex, where she stayed (eventually as Professor) until becoming Principal of Newnham College Cambridge in 1992. She has since been a Professor Emerita in Cambridge. She has written influentially on ethics, political philosophy and the work of Immanuel Kant; her books include *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), *Bounds of Justice* (CUP 2000) and *Autonomy and Trust in Bioethics* (CUP 2002). In addition to her distinguished philosophical career, O'Neill is an active force in public life: she has been a life peer and active cross-bench member of the House of Lords since 1999, was president of the British Academy from 2005 to 2009, and has served on (and chaired) a number of influential public committees and commissions.

by Onora O'Neill

SOMEWHERE IN THE SPACE between realist and relativist accounts of ethics there is said to be a third, distinct possibility. One such position, allegedly both anti-realist and anti-relativist, is John Rawls' 'Kantian Constructivism', first formulated in *A Theory of Justice* and since elaborated in a series of papers.[1] Rawls' critics doubt whether he has found any stable third possibility. On closer inspection, they suspect, every elaboration of the theory depends either on unvindicated transcendent moral claims or on the actual ethical beliefs of some society.

If these suspicions are well-founded, Rawls has not constructed a position that is neither realist nor relativist. Nor have his critics shown the constructivist project impossible. However, they may have shown a great deal about constraints on the wider project by concentrating on the specific difficulties of Rawls' constructivism. I shall discuss Rawls' responses to those critics as a route into an account of a variant constructivism that may be less likely to be absorbed either into moral realism or into relativism.

I. DESIGNS FOR ORIGINAL POSITIONS

Rawls' responses to his critics can be traced in successive formulations of his position. *A Theory of Justice* outlined his well-known method of 'reflective equilibrium' for determining principles of justice. The method seeks coherence between 'our considered moral judgements' and the principles that would be chosen by rational beings whose specific identities and desires have been obscured by a controversially tailored veil of ignorance, which defines a canonical 'original position'. Since coherentist strategies may yield multiple solutions, Rawls offered only *a* theory of justice; there is no claim that reflective equilibrating yields a unique solution. *(TJ,* 50)

His critics detect traces both of relativism and of transcendent moral ideals in this approach. The reference to 'our' considered moral judgement prompted queries about who 'we' are. Can an account of justice that depends even in part on 'our considered moral judgements' avoid contamination by the tenets of corrupt moral traditions, or of traditions that privilege

1.   John Rawls, *A Theory of Justice,* Harvard University Press, Cambridge Mass, 1971; John Rawls, 'The Independence of Moral Theory'. *Proceedings and Addresses of the American Philosophical Association,* 48, 1974-5; John Rawls, 'A Kantian Conception of Equality', *Cambridge Review,* 1975; John Rawls, 'The Basic Structure as Subject', *American Philosophical Quarterly,* 14,1977; John Rawls, 'Kantian Constructivism in Moral Theory', *Journal of Philosophy,* LXXVII, 1980; John Rawls, 'Justice as Fairness: Political not Metaphysical', *Philosophy and Public Affairs,* 14, 1985. Footnotes in the text use obvious abbreviations of these titles: *TJ, BS; KC; JFPM.*

certain sorts of lives and perceptions? Fears that Rawls' thinking cannot escape the claims of entrenched privilege are particularly evident in some discussions of gender justice and of international justice. Here it matters a lot who 'we' are taken to be. Other critics thought that reference to judgements made by idealised agents in an original position uncritically endorsed moral ideals that needed the type of metaphysical vindication that Rawls eschewed. In bracketing 'knowledge of those contingencies which set men at odds' *(TJ,* 19) we deem ideal the principles of beings whose relevance to human affairs looks rather scanty. While Rawls criticised utilitarians for not taking the distinction between persons seriously *(TJ,* 27), some of his critics thought that he did not take the connection between persons seriously. Rawls' responses to both sets of critics can be traced in his successive comments on the nature of constructivism and on its Kantian background.

In *A Theory of Justice* Rawls mainly uses the metaphor of *contract* rather than that of *construction* to characterise his enterprise. Justice as fairness 'carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant' (*TJ,* 11, cf. 3). However, the method of generating principles of justice by reference to choices made in an original position is also said to be 'constructive' because it is a procedure that can settle disputes. Intuitionism is not constructive, because it cannot adjudicate conflicts between intuitions. Rawls offers 'constructive criteria' *(TJ,* 34,39-40,52) by which the considered moral judgements of the status quo may be challenged and revised (*TJ,* 49) without reference to a supposed moral reality. The particular constructive procedures are deemed 'Kantian' because they rely on a more abstract and austere account of the agency and rationality of those who are to 'construct' principles of justice than is used, for example, by utilitarians.

Without doubt important aspects of *A Theory of Justice* are Kantian. However, Rawls departed fundamentally from Kant by relying on a solely instrumental conception of rationality.[2] Justice constructed on this basis must seemingly consist of principles that Kant would have rejected as heteronomous. This danger was to be averted by shrouding the original position in a veil of ignorance. *(TJ,* 252) Those who don't know what they want can't choose self-serving principles. However, instrumental reasoning is wholly at sea unless oriented by some goals. Hence Rawls lifted the veil enough to allow knowledge that certain 'primary' goods—rights and liberties, powers and opportunities, income and wealth *(TJ,* 62)—are universally desired. This filtered access to knowledge of desires is carefully limited to ensure that parties in the

2. *TJ,* 14, 142. For perceptive and diverging views of Rawls' Kantianism see Robert Paul Wolff *Understanding Rawls,* Princeton University Press, 1977 and Stephen Darwall, 'Is There A Kantian Foundation for Rawlsian Justice?' in Gene Blocker and Elizabeth Smith, eds., *Rawls' Theory of Social Justice,* Ohio University Press, 1980.

original position are mutually disinterested. Unless the theory veils the cross-referring structure of human desires—second-orderedness, reactive attitudes to others, altruism, malevolence and envy—an account of human motivation as well as a short-list of primary goods would 'be needed to orient choice in the original position. Detailed motivational premises would introduce controversy and might disable the decision procedures of a revised original position. However, blanket ignorance of second-order desires in turn obscures too much, since it deprives choosers of reasons to care about a future beyond their own lives. Hence Rawls once more lifts the veil to let in a limited assumption about second-order desires: he requires parties to the original position to 'represent continuing lines of claims, as being so to speak deputies for an everlasting moral agent or institution' *(TJ,* 128). The agents in the original position, we discover, are to be thought of as heads or representatives of families who, although generally disinterested in others, care at least about their immediate descendants. Much depends on artful tailoring of the veil of ignorance.

II. ABSTRACT AGENTS AND IDEALISED PERSONS

In *A Theory of Justice* Rawls describes his method as *abstracting* from features of actual human choosing. He aims to argue from 'widely accepted but weak premises' *(TJ,* 18). The 'veil of ignorance is arrived at in a natural way', because we only restrict and do not augment our knowledge of human choosing. We approach a theory of justice minimally by seeing it as part of a theory of rational choice *(TJ,* 16, 47); we hinge nothing on a metaphysically demanding account of the self. What we have is a 'procedural interpretation of Kant's conception of autonomy and the categorical imperative' *(TJ,* 256) that does not depend on the dubious metaphysics of transcendental idealism (cf. *BS165).*

Like the social contract theorists in whose tradition he stands, Rawls has often been criticised for being too abstract. It is not easy to see just why this is a failing. Abstraction, taken strictly, is simply a matter of detaching certain claims from others. Abstract reasoning hinges nothing on the satisfaction or non-satisfaction of predicates from which it abstracts. Rawls hinges nothing on the determinate desires and ideals of particular human beings. Why should this matter? All uses of language must abstract more or less: the most detailed describing cannot abolish the indeterminacy of language. There is no general reason to object to an account of justice that argues from abstract premises to abstract principles. Highly abstract ways of reasoning are often admired (mathematics, physics), even well paid (accountancy, law). Abstract principles are surely needed for reasoning that has broad

scope. Of course, we will also need to apply abstract principles in specific contexts: but that is just as true in law as in ethics, just as true of less abstract, relativised as of the most abstract, supposedly non-relativised ethical principles.

Why then is ethical reasoning so often criticised for being abstract? Is some other difficulty perhaps confused with abstraction? In particular, is there more than abstraction behind the construction of the original position? At first it may seem that Rawls does no more than abstract. The veil of ignorance merely obscures; it only limits claims about agents in the original position. They simply know less than actual human agents. Rawls does not add to their information or desires in any respect. However, the ways in which he abstracts are governed by a certain ideal. The artful tailoring of the veil of ignorance is determined by a highly selective abstraction from actual human choosing, which reflects a certain ideal of the human subject. This ideal is used not just to determine how 'thick' or 'thin' the veil of ignorance should be, but to tear some carefully placed holes in it.

Abstraction, taken strictly, is unavoidable and in itself innocuous. Idealisation is another matter. Objections to supposedly 'abstract' ethical principles and reasoning are often objections to idealisation. The objection is not to reasoning that is detached from certain predicates that are true of the objects discussed, but to reasoning that assumes predicates that are false of them. Reasoning that abstracts from some predicate makes claims that do not hinge on the objects to which it is applied satisfying that predicate. Reasoning that idealises makes claims that apply only to objects that live up to a certain ideal.

The veil of ignorance described in *A Theory of Justice* was tailored to hide the interlocking structure of desires and attitudes that is typical of human agents. Once the social relations between agents were masked it could seem plausible to assign to each desires for a uniform short-list of primary goods, and to build a determinate ideal of mutual independence into a conception of justice. This ideal is not met by any human agents. It isn't only deficient and backward human agents whose choosing would be misrepresented by these ideal agents of construction. The construction assumes a mutual independence of persons and their desires that is false of all human beings. Such independence is as much an idealisation of human social relations as an assumption of generalised altruism would be.[3]

Idealisations have no doubt many theoretical advantages: above all they allow us to construct models that can readily be manipulated. However, they may fail to apply to any significant domain of human choosing.

3. Compare Arthur Ripstein, 'Foundationalism in Political Theory', *Philosophy and Public Affairs*, 16, 1987. Ripstein argues that Hobbes' accounts both of instrumental rationality and of mutual indifference are not abstractions from a state of nature, but ungrounded idealisations.

This is sometimes defended by construing ideals not as abstractions but as 'simplifications'. This too is strictly inaccurate. A theory simplifies if it either leaves things out (i.e. abstracts) or smooths out variations. If it incorporates predicates that are not even approximately true of the agents to whom the model is supposed to apply it does not simplify. If idealisations do not 'simplify' the descriptions that are true of actual agents, then they are not innocuous ways of extending the scope of reasoning. They covertly endorse 'enhanced' versions of specific human characteristics and capacities. Idealisation masquerading as abstraction produces theories that may appear to apply widely, but in fact covertly exclude from their scope those who don't match a certain ideal. They privilege certain sorts of human agent and life by presenting their specific characteristics as universal ideals.

III. TOWARDS KANTIAN CONSTRUCTIVISM

In the papers Rawls published during the seventies the role of a certain ideal of the person was fully acknowledged. In *A Theory of Justice* the principles of justice defined, but were said not to presuppose, an ideal of the person. *(TJ, 260-61)* Later papers took increasing account of the complex structure of human desires. Here Rawls argued that an ideal of the moral person played an essential part in determining the principles of justice. The 1980 Dewey Lectures no longer see primary goods just as essential means to whatever ends human beings actually have. They are rather 'generally necessary as social conditions and all-purpose means to enable human beings to exercise their moral powers' *(KC, 526)*. These 'moral powers' are capacities to develop a sense of justice and a conception of the good; they are the motivational analogues, and antecedents, of Rawls' two principles of justice. These powers define a certain ideal of the moral person and thereby the highest-order interests of actual persons. This ideal is now built into the original position, whose agents of construction are 'moved solely by the highest-order interests in their moral powers and by their concern to advance their determinate but unknown final ends' *(KC, 528, cf. 547; 568)*. This revised account of the original position explicitly bases principles of justice not on a conception of the person that abstracts from the diversity of human agents, but on one that idealises a certain sort of agent.

The ideal which Rawls elaborates is 'the Kantian ideal of the person'. By stressing that it is an *ideal* of the person that informs the original position, Rawls avoids endorsing an account, let alone the suspect Kantian account, of the metaphysics of the self. Procedural Kantianism endorses only (a version of) the ideal of the inde-

pendent and autonomous character which we associate with Kant. In this way obscure and panicky metaphysics are to be avoided.

Yet ideals too need vindication. How can Rawls do this if he does not offer a metaphysics of the self or of the person? Some critics of Kantian constructivism believed that he had committed himself to a task that was impossible within the anti-realist constraints that he sets on his moral theory.[4] Rawls' basic assumptions about agency and rationality set only weak feasibility constraints on ideals of the person. Differing ideals of the person could form part of other constructive procedures that would generate other conceptions of justice. Rawls indeed acknowledges the possibility of variant constructivisms *(KC, 535;* cf. *TJ,* 17). He therefore needs to vindicate a specific ideal of the person. Either he must show that the Kantian ideal is a 'uniquely plausible ideal of the person'[5] or he must provide some other reason for singling it out.

'Justice as Fairness: Political not Metaphysical' provides that other reason. Here Rawls still avoids 'claims to universal truth about the essential nature and identity of persons' *(JFPM,* 223). 'Justice as fairness', he now argues, 'is a political conception of justice because it starts from within a certain political tradition' *(JFPM,* 225). The ideal of the person on which his argument rests is not that of the abstract individual (as certain critics had supposed), but that of persons as citizens of a modern democratic polity, who (while they may disagree about the good), accept the original position as a 'device of representation' *(JFPM,* 236) that accurately captures their ideal of a fair system of cooperation between citizens who so disagree. Far from deriving a justification of democratic citizenship from metaphysical foundations, Rawls invites us to read *A Theory of Justice* as a recursive vindication of those deep principles of justice 'we' would discover in drawing on 'our' underlying conceptions of free and equal citizenship. This vindication of justice does not address others who, unlike 'us', do not start with such ideals of citizenship; it has nothing to say to those others. It is 'our' ideal, and 'our' justice. Worries about Rawls' relativism come flooding back.

In a way these worries are intensified, for it is not just by way of equilibrating theory with 'our considered moral judgements', but in the very formulation of the most abstract theoretical principles that Rawls now appeals to the judgements of 'our' tradition. Kantian constructivism, it seems, claims only to offer a coherent articulation of the outlook of modern liberal societies. The Kantian ideal of the person is socially embedded, and anti-relativism is not attainable. We are offered a coher-

4. David O. Brink, 'Rawlsian Constructivism in Moral Theory', *Canadian Journal of Philosophy,* 17, 1987, 71-90, esp. p. 73.

5. *Ibid,* p. 90.

ent articulation of the deep moral commitments of 'our' society. With hindsight many of Rawls' earlier writings can be seen to acknowledge this derivation of justice.

IV. TOWARDS A MORE KANTIAN CONSTRUCTIVISM

Was there any other possibility? Could Rawls have avoided idealising premises and relied on a genuinely abstract yet non-idealising account of agency and rationality? If so, what would he have relied on?

If we stand back from the entire project we can see that many of the difficulties of formulating a 'Kantian' theory of justice arose from uncertainty about the degree to which principles of construction should rely on and reflect desires. Once Rawls had committed himself to a merely instrumental conception of rationality, all desires would be reflected in the outcomes of construction, unless specifically bracketed. The veil of ignorance had to be spread to avoid heteronomy. The veil then had to be breached selectively to avoid complete indeterminacy. A specific ideal of the person had to be invoked to explain just how the veil should be tailored. Finally 'our' tradition was invoked to vindicate this ideal without having to establish metaphysical claims about agency. In short, some of the *least* Kantian features of Rawls' constructivism produced a train of difficulties.

Could a more nearly Kantian constructivism do better? It could surely avoid the need to vindicate selective bracketing of desires, by detaching principles of justice from all claims about desires. However, a more Kantian constructivism would be pointless if it merely rehearsed the supposed empty formalism of Kant's ethics, and so fell foul of Mill's old charge of failing grotesquely to derive any actual principles of duty. Can a construction do any work if it draws neither on decontextualized and unvindicated accounts of ideal rationality and independence, nor on the moral ideals of a specific culture?

Rather than tackle the question in general form, I shall sketch the outlines of another construction. Constructivism might begin simply by abstracting from the circumstances of justice, meagrely construed. The problem of justice arises only for a plurality of at least potentially interacting agents. It does not arise where there is no plurality, or no genuine plurality of agents, hence no potential for conflict. (The action of 'agents' in such a degenerate plurality might be automatically coordinated, perhaps by instinct or by a pre-established harmony.) Nor does it arise among agents who cannot interact—castaways, isolates, the men and women of Rousseau's earliest state of nature are outside circumstances of justice.

Some assumptions about the agents of such a plurality are needed if any construction is to be possible. If idealisation is to be avoided, these assumptions must only abstract. They must not smuggle in reference to unvindicated moral ideals. A more Kantian constuctivism (perhaps not the one everybody would attribute to Kant[6]) must then start from the *least determinate* conceptions both of the rationality and of the mutual independence of agents.

A meagre and indeterminate view of rationality might credit agents only with the capacity to understand and follow *some* form of social life, and with a commitment to seek *some* means to any ends (desired or otherwise) to which they are committed. (This weak view of instrumental rationality is non-committal about the efficient pursuit of ends: efficiency comes into play only if some metric of 'costs' and 'benefits' together with procedures for decomposing action into 'options' is assumed. A construction premised on efficient pursuit of ends covertly privileges specific moral and social ideals.)

A meagre and indeterminate view of the identity and of the mutual independence of agents can assume only that agents have capacities for varying sorts and degrees of dependence and interdependence. A complete erosion of capacities for independent action destroys plurality and with it the context of justice; a complete erosion of dependence privileges an ideal of the person whose relevance to human life is wholly unestablished. Natural persons are always artificial too, in many ways. Their beliefs and desires and their very identities can interlock in many ways, which only approximate to varying ideals of the person. Constructivism which does not privilege any one ideal of the person cannot be premised on a fixed account of the forms of rationality or the degree of mutual independence of agents. It must seek principles for agents who are numerous, not ideally rational and not ideally independent of one another.

As sketched this blueprint seems too indefinite to guide any construction. It surely cannot guide an answer to the *hypothetical* question 'What principles *would* a plurality of agents, with minimal rationality and indeterminate capacities for independence, choose to live by?'. Materials for answering hypothetical questions are lacking. However, the blueprint may permit a beginning of an answer to a *modal* question such as 'What principles *can* a plurality of agents of minimal rationality and indeterminate capacities for mutual independence live by?'. No plurality can choose to live by principles that aim to destroy, undercut or erode the agency (of whatever determinate shape) of some of its members.[7] Those who become victims of action on such principles not merely

6. This may seem gross understatement. Standard readings of Kant's moral philosophy construe his starting point as metaphysically exorbitant rather than modestly abstract. The interpretive issues hinge on determining the starting point of the critical enterprise. It is at least arguable that if we are to make any sense of the Kantian enterprise we must take the notion of critique of reason as the most fundamental, and that in doing so we may find that Kant's underlying moves, which govern his account of philosophical method in the Doctrine of Method of the *Critique of Pure Reason,* constitute a constructivist vindication of reason.

7. This throws a heavy weight on the question of membership. Who counts in universalizability tests? An adequate answer for present purposes is that we cannot exclude from membership those with whom we interact and on whose rationality we rely. Distant others on whose abilities to translate, negotiate and trade we presume cannot be excluded; extraterrestrials can be (for the time being). Nor can we anticipate the loss of capacities for agency by victims of violence in order to vindicate inflicting that loss.

do not act on their oppressor's principles: they *cannot* do so. Victims cannot share the principles on which others destroy or limit their very capacities to act on principles.[8]

V. ENTITLEMENTS OR OBLIGATIONS?

If we are to move from this very indeterminate claim about the principles that a plurality of agents, of whom we assume only minimal rationality and indeterminate mutual dependence, could live by, we need to take some view of the task of those principles. In *A Theory of Justice* Rawls noted that constructivism might undertake various tasks: 'the contractarian idea can be extended to the choice of more or less an entire ethical system . . . including principles for all the virtues' *(TJ,* 17). His own more limited aim was to construct principles of justice that would determine the basic structure of a society, so constraining the actions of institutions and of individuals. What he actually argues for are specifically principles of *entitlement* rather than of *obligation.* His first principle of justice assigns to each 'an equal right to the most extensive basic liberty compatible with like liberty for all' (*TJ,* 60). The second principle's main demand, as finally stated, is that social and economic inequalities be arranged to the greatest benefit of the least advantaged (*TJ,* 302). A battery of objections has been fired at these claims. In particular, the priority assigned to liberty has been diagnosed as privileging the very ideal of mutual independence that has already been queried. For present purposes the details of these debates do not matter. However, Rawls' focus on entitlement rather than on obligation is significant. Constructing a set of entitlements is not the same as constructing principles of obligation.

This is perhaps surprising. Within an account of justice it may seem unimportant whether we adopt the perspective of agents and their obligations, or of recipients and their entitlements. The set of obligations and the set of entitlements will presumably be reciprocally defined. (This would not be the case if we were considering a wider range of obligations, some of them 'imperfect' obligations to whose performance nobody was entitled.) However, the perspective of recipience and entitlement has other difficulties that obstruct the project of construction.

A constructivist approach that stays with the perspective of entitlement that Rawls prefers, but asks the Kantian, modal question, has to look for a set of entitlements that can consistently be held by all. There are many compossible sets of entitlements, so a further move is needed to identify the just set. It is tempting to think that this move can be made simply by seeking

8. Colluding victims are not a counterexample. If collusion itself is coerced they are not truly willing; if it is not they are not truly victims. Although the principle is clear, cases are often highly ambiguous: witness the Patty Hearst trial.

the *maximal* set of entitlements that can consistently be held by all. This move would parallel the two moves Rawls makes to determine first which system of liberties and then which social and economic inequalities are just. Each of Rawls' principles purports to identify a *maximal* set of entitlements. We can find maxima only where there is some metric. Unfortunately liberty has no metric, and social and economic arrangements have one only on utilitarian premises that Rawls skirts and that a more Kantian constructivism would reject. By detaching justice from the desires either of idealised or of actual agents we discard the framework needed to make sense of metric and maximising notions. The notion of a maximal set of entitlements (whether liberties or social and economic arrangements, or the two taken together) is indeterminate.

However, principles of justice could be fixed by constructing principles of obligation rather than of entitlement. In the tradition of the social contract theory, but not in its contemporary descendants, principles of justice define obligations rather than entitlements. A return to this perspective is, I believe, required for a non-idealising constructivism because obligations of justice, unlike entitlements, can be constructed without assuming a metric either for liberty or more generally for actions. At a later stage of the argument a constructivist approach to justice which successfully identifies principles of obligation by using modal arguments may also identify the entitlements that are the reciprocals of these obligations.

VI. A CONSTRUCTION OF JUSTICE

Everything then hinges on constructing principles of justice without presupposing a determinate ideal of the person and without privileging the perspective of entitlements and rights. The core of any such construction is the thought that there are certain constraints on the principles of action that could be adopted by all of a plurality of potentially interacting agents of whom we assume only minimal rationality and indeterminate mutual independence. Principles which cannot be acted on by all must be rejected by any plurality for whom the problem of justice arises.

This may seem too meagre a basis for justice. One classical objection to Kantian formalism has been that *any* internally coherent principle for individual action is universalizable: what is open to any arbitrarily chosen agent is open to all, apart from uninteresting exceptions that refer to unique performances, competitive successes and the like. This formalistic interpretation of universalizability ignores the constraint posed by the fact that we are considering the case of a plurality of *potentially*

interacting beings, that is of beings who share a world. Any principle of action that is adopted by all members of such pluralities affects the world that they share and becomes a background condition of their action. This is why certain principles of action which can coherently be adopted by some cannot be coherently adopted by all. Justice, taken in the traditional, minimal, formal sense of like requirements for like cases, will then require that those principles be rejected.

Examples of principles that cannot be universalized can illustrate the point. A principle of deception, which undermines trust, would if universally adopted, destroy all trust and so make all projects of deception impossible. Selective deception is on the cards: universal deception is not. A principle of coercion, whose enactment destroys or undercuts the agency and will of at least some others for at least some time, cannot be universally followed. Those who are at a given time the victims of coercion cannot act, so cannot make coercion their own principle. Equally, action on a principle of violence damages the agency of some, so cannot be universally acted on. Put quite generally, principles of action that hinge on victimising some, whether by destroying, paralysing or undercutting their capacities for agency for at least some time and in some ways, can be adopted by some but cannot be universally adopted.

VII. FROM PRINCIPLES TO JUDGEMENTS

To keep matters under control let us assume only that justice demands (at least) that action and institutions not be based on principles of victimisation (deception, coercion, violence).[9] Still, it may seem, we are far from showing what justice demands, since we do not know what non-deceit or non-coercion would demand in specific circumstances. These guidelines are highly indeterminate. Even if the most formalistic objection to 'Kantian' formalism fails, have we not paid the classic price of abstraction, reaching highly abstract principles which do not tell us what to do in specific contexts?

Priciples are always to some extent abstract; but they are not the whole of practical reasoning. They must always be applied in ways that take account of actual context: and they never determine their own applications. Even the culturally specific principles that relativists favour do not determine their own applications. So a fuller account of practical reasoning must say *something* about processes of judgement and deliberation; but it cannot say everything. Ethical principles are not algorithms.

How much should be said? Should we expect the most abstract principles of justice to entail a specific

9. There may be other principles of justice: the constructive procedure sketched here does not determine a complete set of principles of justice. This is simply a corollary of detachment from metric assumptions. Because the parallel modal construction of entitlements presents just entitlements as the mutually limiting compossible components of a maximal set, it can identify all just entitlements if it identifies any. Equally if it cannot identify all, it identifies none. A modal construction of obligations establishes each principle by showing why its rejection cannot be universally adopted. Hence obligations can be identified seriatim, and completeness is not guaranteed; indeed it may not even make sense to speak of a *complete* set of principles of action.

code, for example one to which legislators could appeal? Some people still clearly expect principles alone to give very detailed guidance. For example, there is a luxuriant and sometimes fierce philosophical debate on coercion. This debate aims to distinguish threats (and possibly offers) that coerce from those that do not. If non-coercion is a requirement of justice, should we worry that this debate remains inconclusive? I would suggest that there is no reason to expect the issues to be resolved *unless we agree on a determinate conception or ideal of the person.* Coercion is a matter of force or threat, and what constitutes threat must vary with the vulnerabilities of those who are threatened. Vulnerability depends on many things, including the forms of rationality and of dependence and independence that particular agents have at particular times. Coercers know very well that successful threats take account of victims' specific vulnerabilities. It is unreasonable to look for a set of necessary and sufficient conditions of coercion that will apply quite generally unless we can vindicate a specific ideal of the independence of persons, which will provide a standard for distinguishing those who weakly succumb to mere gestures of threat from those who forgivably yield to overwhelming duress. Parallel comments could be made on non-deception and non-violence.

What then can be said about the move from an abstract principle to judgements about specific institutions or actions if we lack a determinate ideal of the person? Is there any way to operationalise the idea of rejecting unsharable principles, without subordinating it to the categories and views of the status quo? Even if we have found principles whose *vindication* does not depend on accepted moral views, we do not escape relativism if their *application* unavoidably endorses the categories and concerns of the powers that be, rather than subjecting them to scrutiny.

To do this we must rely not merely on the abstract principles that could be accepted (or rejected) by any plurality of minimally rational agents of indeterminate mutual independence. We must rely on specific interpretations of these principles that can be accepted (or rejected) by those actually involved. Only action and policies that guarantee the refusability by actual agents of offers and involvements that others propose can ensure consent that is not merely nominal but legitimating.

The appeal to consent here is neither to the hypothetical consent of the ideally rational and independent nor to an actual consent that might reflect oppression. (Those routes lead back to realism and to relativism.) Rather the appeal is to the *possible consent* of *actual agents*. The criterion for this is not that consent is os-

tensibly given (that might reflect false consciousness or duress) but that any arrangements or offers could have been refused or renegotiated. If we are to be sure that a principle could have been shared even by those on whom it bears hard, we need to be sure that they could (even if ignorant and weak) have refused or renegotiated the roles or tasks that action on that principle imposed on them. Neither the apparent consent of the vulnerable nor the hypothetical consent of the glitteringly self-sufficient legitimates. Equally the absence of expressed consent in those who have opportunity and capacities to refuse does not signal injustice.

Thinking in this way about applying principles of justice we can see that *it demands more, not less, to be just to the vulnerable* and that *genuine, legitimating consent is undermined by the very institutions which most readily secure an appearance of consent*. The vulnerable are simply easier to deceive and to victimise than the strong. By contrast both idealised and relativised accounts of justice tend to conceal the fact that justice to the weak demands more than justice to the strong. Idealised accounts of justice tend to ignore actual vulnerabilities and relativised accounts to legitimize them.

This line of thought can be applied to institutional as well as individual injustice. When relations between agents are ones of structured dependence, it is hard or impossible for agents to refuse an apparent consent to arrangements which structure their lives and identities. The weak risk too much by dissenting unless institutions are structured to secure the option of refusal: and when this option is secured they will no longer be so weak.

Put generally the point is that in applying abstract, non-idealising principles we have to take account not indeed of the actual beliefs, ideals or categories of others, all of which may reflect unjust traditions, but of others' actual capacities and opportunities to act—and of their incapacities and lack of opportunities. This move does not lead back to relativism; no principle is endorsed simply because currently accepted. A more nearly Kantian constructivism would use modal notions to identify principles, but indicative ones to apply them. The principles of justice hold for any possible plurality: for they demand only the rejection of principles that cannot be shared by all members of a plurality. The determination of justice for actual situations is regulated but not entailed by these principles. The feature of actual situations that must be taken into account in judgements of justice is in the first place the security or vulnerability of agents that allows agents to dissent from the arrangements which affect their lives and whose absence compromises any ostensible 'consent'.

Rawlsian constructivism has ended up on an uncomfortable knife edge, and teeters between idealising and relativized conceptions of ethics. The idealised versions demand proofs of a moral or metaphysical reality Rawls does not discern: the relativized readings can only offer internal critique of the justice of modern liberal societies.

Several features distance the alternative constructivism sketched here from Rawls' pioneering project. First, idealized accounts of agents, their rationality and their mutual independence are explicitly rejected: there is no appeal, however oblique, to transcendent moral claims. Second, abstraction from the determinate desires of agents is complete: hence no special ingenuity is needed to avoid either heteronomy or cruder forms of relativism. Third, and as a consequence, it is impossible to answer questions about the hypothetical choices of abstract agents. Fourth, the construction therefore has to fall back on modal questions about the possible choices of abstract agents, and construct an answer to the question 'What principles must a plurality of abstractly characterised agents reject?'. Finally, rejection of non-universalizable principles can guide action by requiring that we ensure that the agents actually affected, with their particular identities and vulnerabilities, can genuinely choose or refuse those principles. If this sketch can be filled out, there is at least some space between realism and relativism.

O'NEILL'S 'CONSTRUCTIVISMS IN ETHICS'

Commentary by Connie S. Rosati

O'NEILL'S 'CONSTRUCTIVISMS IN ETHICS'

Commentary by Connie S. Rosati

CONNIE S. ROSATI is
Associate Professor
of Philosophy at the
University of Arizona.
Her research interests are
principally in metaethics
and the philosophy of
law. She is the author of
"Persons, Perspectives, and
Full Information Accounts
of the Good" (*Ethics*
1995), "Personal Good"
(in *Metaethics After Moore*
2006), "Objectivism and
Relational Good" (*Social
Philosophy & Policy*
2008), "The Story of a
Life" (*Social Philosophy
& Policy* 2013), and
"The Makropulos Case
Revisited: Reflections on
Immortality and Agency"
(*Oxford Handbook of
Philosophy of Death* 2013).

Commentary by Connie S. Rosati

IN HER 1988 PRESIDENTIAL ADDRESS TO THE ARISTOTELIAN SOCIETY, Onora O'Neill discusses problems critics have raised for John Rawls' version of constructivism, and she sketches a "more Kantian" constructivism intended to overcome these difficulties.[1] The challenge, as O'Neill sees it, is to devise a form of constructivism that (A) escapes objections to Kant's formalism, (B) does not depend, as critics allege Rawls' view does, on either "unvindicated transcendent moral claims" or on the "actual ethical beliefs of some society," and (C) is the genuinely distinct alternative to both realism and relativism that critics allege Rawls' view is not. (246) The key to escaping objections to Kant's formalism, according to O'Neill, lies in "using modal notions to identify principles, but indicative ones to apply them."(258) The key to avoiding alleged problems for Rawlsian constructivism, and so to finding a middle ground between realism and relativism, lies in avoiding appeal to transcendent moral claims, controversial ideals of the person, and actual moral beliefs of particular societies, while still abstracting from the determinate desires of agents. I herein examine O'Neill's intriguing address. As I shall explain, it is doubtful that she has succeeded in providing a compelling alternative to Rawlsian constructivism, one that is, as she conjectures, "less likely to be absorbed either into moral realism or into relativism."(246) In fact, the problems with her sketch go some way to vindicating a more Rawlsian approach.

I. RAWLS AND HIS CRITICS

Rawls' critics contend that he has failed to find a stable third possibility, between realism and relativism, O'Neill observes. But they "have not shown the constructivist project impossible."(246) O'Neill suggests that, in fact, Rawls' critics have taught us about constraints on the broader constructivist project by means of the difficulties they raise for Rawls' particular version of constructivism. She focuses on Rawls' responses to these critics "as a route into an account of a variant constructivism."(246)

After surveying how Rawls' view developed over the years, partly in response to problems raised by his critics, O'Neill concludes that "some of the *least* Kantian

features of Rawls' constructivism produced a train of difficulties" (252). Among the difficulties she identifies for Rawls' constructivism, are its offering an idealization of human social relations that "assumes a mutual independence of persons and their desires that is false of all human beings,"(249) its reliance on a purely instrumental conception of rationality, (247-8) and its tailoring of the original position in ways that reflect a certain "ideal of the human subject."(249) Idealizations, O'Neill complains, "may fail to apply to any significant domain of human choosing" and may covertly "privilege certain sorts of human agent and life by presenting their characteristics as universal ideals."(250)

Rawls is clear, in *A Theory of Justice*, about relying on idealization of the parties and of the choice situation for selection of principles of justice.[2] But it is only in later papers, and especially in his 1980 Dewey Lectures, "Kantian Constructivism in Moral Theory," that Rawls explicitly acknowledges relying on an ideal of the moral person.[3] Critics complained that because there are other ideals of the person, and so other possible conceptions of justice, he needed to establish that his ideal of the person was most plausible or that there is reason for accepting it over others. (250-1) But it would seem that if his ideal of the person were most plausible, then his constructivism would collapse into realism, and if it were not, his constructivism would collapse into relativism.

Indeed, the latter seems to be precisely what happened, O'Neill suggests, because of the tack Rawls took for dealing with this criticism. Rather than argue that his ideal of the person is most plausible, he argues, in "Justice as Fairness: Political, not Metaphysical," that it is the ideal of "persons as citizens of a modern democratic polity."[4] Rawls offers a vindication of principles of justice that apply only to us members of modern liberal societies, and so one that need not be accepted by those who favor a different ideal.

O'Neill diagnoses the problems for Rawls' view as follows:

> If we stand back from the entire project we can see that many of the difficulties of formulating a 'Kantian' theory of justice arose from uncertainty about the degree to which principles of construction should rely on and reflect desires. Once Rawls had committed himself to a merely instrumental conception of rationality, all desires would be reflected in the outcomes of construction, unless specifically bracketed. The veil of ignorance had to be spread to avoid heteronomy. The veil then had to be breached selectively to avoid complete indeterminacy. A specific ideal of the person had to be invoked to explain just how the veil should be tailored. Finally 'our' tradition was invoked to vindicate this ideal without having to estab-

2. Rawls (1971: sec. 11, 60).

3. Rawls (1980: 535)

4. Rawls (1985: 7).

lish metaphysical claims about agency. In short, some of the *least* Kantian features of Rawls' constructivism produced a train of difficulties. (252)

O'Neill recognizes the need to abstract from individuals' actual desires. What she questions is the need for idealization.

II. TOWARD AN ALTERNATIVE CONSTRUCTIVISM

O'Neill asks, "Was there any other possibility? Could Rawls have avoided idealizing premises and relied on a genuinely abstract yet nonidealizing account of agency and rationality?" (252) She goes on to present the outlines of a "more nearly Kantian constructivism" along these lines.

According to O'Neill, in order to arrive at a more viable, and more nearly Kantian, constructivism, we need to shift away from the Rawlsian model as follows:

1. Focus on the idea of a *plurality of interdependent agents*, rather than on a "group" of mutually disinterested agents, so as to match the circumstances of justice.

2. Employ a conception of agents as *indeterminately rational and interdependent*, so as to avoid relying on an ideal of the person.

3. Alter the task of principles of justice to setting *obligations* rather than entitlements, so as to avoid problems stemming from the indeterminacy of the notion of a maximal set of entitlements. (254-5)

Consider the first shift. Rawls assumes mutually disinterested agents choosing principles of justice from behind the veil of ignorance. But the problem of justice arises only for a plurality of potentially interacting agents; it arises, that is, only where a potential for conflict exists. So our construction should focus instead on a *plurality of potentially interacting agents*.

Now, the second shift. Rawls' characterization of agents rests on a (controversial) ideal of the person that assumes a determinate degree of mutual independence and that involves a determinate conception of their rationality, namely, the instrumental conception. Instead, characterize agents in a way that eschews idealization, so as to avoid illicit use of "unvindicated moral ideals." (253) If constructivism is to take a more indeterminate view of the identity and mutual independence of agents, it can "assume only that agents have capacities for varying sorts and degrees of dependence and independence." (253) If agents are assumed to have no capacity for inde-

pendent action, then there is no plurality, and so no set of circumstances requiring principles of justice. If agents are assumed to have no dependence on one another, this "privileges an ideal of the person whose relevance to human life is wholly unestablished." (253) Assume, then, the least determinate conception of the mutual independence of agents.

Assume also the least determinate conception of the rationality of agents. (252-3) If constructivism is to avoid privileging a particular ideal of the person, then it cannot rest on a particular view of rationality any more than it can rest on a particular view of the mutual dependence or independence of agents. (253) In adopting a more indeterminate view of their rationality, we might treat agents as having the capacity only to "understand and follow *some* form of social life," and a commitment only "to seek *some* means to any ends (desires or otherwise) to which they are committed."(9-10) The weak form of instrumental rationality that a more Kantian constructivism would suppose "is non-committal about the efficient pursuit of ends," thereby avoiding privileging particular moral and social ideals.[5] (253)

Constructivism, in short, "must seek principles for agents who are numerous, not ideally rational and not ideally independent of one another." (253) O'Neill acknowledges that this sketch—and it is just a sketch—will not enable us to answer the *hypothetical* question "What principles *would* a plurality of agents, with minimal rationality and indeterminate capacities for independence, choose to live by?" But it might, she thinks, enable us to begin to answer a *modal* question, such as "What principles *can* a plurality of agents of minimal rationality and indeterminate capacities for mutual independence live by?" (253)

In distinguishing between hypothetical and modal questions, O'Neill seems to draw two distinctions—between what agents *would* do and what it is *possible* for them to do, and between what agents would *agree to* and what they can *live by*. What turns on these distinctions? And how does her sketch permit the beginnings of an answer to the modal question? O'Neill isn't explicit about this, but her discussion suggests that the modal question concerns a *possible consensus* of a plurality of agents. "No plurality can choose to live by principles that aim to destroy, undercut or erode the agency (of whatever determinate shape) of some of its members." Why is that? Because victims of such principles cannot share those principles, they "cannot share the principles on which others destroy or limit their very capacities to act on those principles."[6] (254)

As for the third shift, Rawls sets for parties to the

5. O'Neill doesn't make clear what moral and social ideals are thereby privileged, and one might well doubt her claims.

6. O'Neill doesn't explain how we are to understand this claim. I return to this problem shortly.

original position the task of choosing principles of justice that are really principles of entitlement. This is manifested, in particular, in the first principle of justice, the principle of equal liberty.[7] Once we have switched to the modal question, the emphasis on entitlements creates problems for constructivism, for the constructivist must now "look for a set of entitlements that can consistently be held by all." (254) It won't do to try to find the "maximal set" of entitlements, O'Neill claims, because for this, we need a metric. But "liberty has no metric, and social and economic arrangements have one only on utilitarian premises that Rawls skirts and that a more Kantian constructivism would reject. By detaching justice from the desires either of idealized or of actual agents we discard the framework needed to make sense of metric and maximizing notions." (255) We need to return, she suggests, to the social contract tradition, in which principle of justice set out obligations rather than entitlements. Obligations of justice, in contrast to entitlements, can be constructed without the need for a metric. Identifying principles of obligation on the basis of "modal arguments" may help us to identify entitlements in turn. (255)

III. A COMPELLING ALTERNATIVE?

Whatever the problems may be for Rawls' version of constructivism, I am not persuaded that O'Neill's sketch offers a compelling alternative. It seems to me that it does not avoid the problems of Kantian formalism, does not successfully chart a course between realism and relativism, and avoids the problem of relying on an ideal of the person only at a serious cost.

First, a word about O'Neill's rejection of idealization. O'Neill has not, I think, offered good reasons for favoring abstraction over idealization, and in fact, some of the difficulties her sketch faces seem to arise precisely because she eschews idealization. Now, I take it to be uncontroversial that constructivism "must seek principles for agents who are numerous, not ideally rational and not ideally independent of one another." (253) In fact, Rawls' constructivism sets out to do precisely that. The real issue between O'Neill and Rawls is not, then, as to *what* constructivism must do but as to *how* it must do it if we are to arrive at plausible principles of justice. To see what is at issue between them, consider that "construction" of principles of justice is done from some standpoint or other.[8] What is at issue between O'Neill and Rawls might be characterized in terms of a difference as to the proper standpoint. Is the standpoint from which to seek principles for agents who are numerous, not ideally rational, and not ideally independent of one another the standpoint of *those agents just*

7. Rawls (1971: sec. 11, 60).

8. vSee, e.g., Shafer-Landau (2003: 15), for a characterization of constructivism as involving a standpoint, or stance dependence. See also Street (2010: 364).

as such? Rawls, in taking up and attempting to answer the hypothetical question, answers "no," and for what we can see to be excellent reasons. After all, the process of selecting principles would be affected by differences in information and rationality, with different agents more or less well informed and more or less rational. It would also be affected by motivational differences, with different agents more or less self-interested and more or less altruistic. Given the potential for manipulation and distortion that these differences create, we lack adequate reason to think any principles resulting from nonidealized construction would treat everyone fairly. For these reasons, Rawls distinguishes between the standpoint of actual agents—and even of O'Neill's abstract agents—and the standpoint of idealized agents, who choose for the circumstances in which they are their nonideal selves.

Consider an analogy to informed desire accounts of an individual's non-moral good, such as Peter Railton's.[9] According to the latter accounts, what is good for a person, P, is what P would want her actual self to want, were she fully informed and rational. The idealization would be problematic, of course, if it were to confuse the good of the person who is fully informed and rational with the good of her actual self.[10] But as these accounts make clear, the idealized person, or P+, asks not what to want for herself as idealized, but what to want for her actual, non-ideal self in her non-ideal circumstances. Rawls similarly distinguishes between ideal and actual world, between evaluating and evaluated world.

What this means is that it doesn't matter, at least not for the reasons O'Neill gives, whether constructivism treats ideal agents in ideal conditions as mutually disinterested rather than interacting, as instrumentally rational rather than minimally rational.[11] What does matter is that it treats ideal agents as choosing principles for the world in which they *will* be interacting and only minimally rational.

To be sure, idealization of the kind employed in Rawlsian constructivism is not unproblematic, even if it does not have all of the problems O'Neill seems to find with it. But turning away from the hypothetical question to the modal question only leads us to different problems. If we assume agents have only minimal rationality—the capacity to seek *some* means to their ends—then it might well be possible for agents to share principles that are destructive to some. With only minimal rationality, after all, some might not appreciate the implications of principles; they might do poorly at choosing means to their ends; their ends might be irrational or self-destructive. As far as I can tell, nothing in O'Neill's sketch precludes such a possibility. O'Neill insists that the prospective

9. Railton (1986a, 1986b).

10. For discussion of the problem and how to avoid the conditional fallacy, see, e.g. Pettit and Smith (1993).

11. More precisely, if it matters, it is not because the idealization involves assumptions that are "false of all actual human beings."

victims of such principles *cannot* share them, that "no plurality can choose to live by" such principles. But it isn't clear how we are to understand these modal claims. To make it plausible that prospective victims *cannot* share these ends, it must either be impossible *conceptually* to share them, or it must be impossible *rationally* to share them. We would need more argument than O'Neill offers, however, to think that conceptual impossibility will rule out much as unsharable; so appealing to conceptual impossibility will not get us far. But agents would have to be more than indeterminately rational to rule out much as rationally unsharable.

Without idealization, then, the modal question arguably resists much in the way of an answer, and so much in the way of a compelling answer. But this problem seems to be much the same as the problem of Kantian formalism. (255-6) As O'Neill sums up the latter problem, "One classical objection to Kantian formalism has been that any internally coherent principle for individual action is universalizable: what is open to any arbitrarily chosen agent is open to all, apart from uninteresting exceptions that refer to unique performances, competitive successes and the like." (255) O'Neill insists that this interpretation of universalizability

> ignores the constraint posed by the fact that we are considering the case of a plurality of potentially interacting beings….Any principle of action that is adopted by all members of such pluralities affects the world that they share and becomes a background condition of their action. This is why certain principles of action which can coherently be adopted by some cannot be coherently adopted by all. (255-6)

She gives as examples a principle of deception and a principle of coercion. Now, I'm inclined to agree that the objection to Kant was never quite accurate. But even if a Kantian Constructivism along the lines that O'Neill envisions could rule out the most extreme principles—and that depends again on how to understand the modal claim that they *cannot* be universally adopted—it would seem still to leave us with a great deal of indeterminacy.

The indeterminacy I have in mind is not the kind O'Neill discusses, namely, that Kantian Constructivism does not tell us what to do in specific situations. The problem of indeterminacy is, rather, that it has insufficient constraints to select principles of justice or to markedly narrow the field. We may learn what principles agents *can't* live by, but that leaves open the possibility of many conflicting principles that they *can* live by. If this is right, then we would seem to be back to the worry that constructivism collapses into relativism, and so does not represent a genuine alternative to either realism or relativism.

It is worth considering O'Neill's response, though, to the other kind of indeterminacy. O'Neill observes that principles do not determine their own applications but must be applied to actual contexts. Regarding whether issues raised by efforts to apply principles, such as what distinguishes threats and offers from coercion, can be resolved, she remarks, "I would suggest that there is no reason to expect the issues to be resolved *unless we agree on a determinate conception or ideal of the person*….It is unreasonable to look for a set of necessary and sufficient conditions of coercion that will apply quite generally unless we can vindicate a specific ideal of the independence of persons…" (257) So what are we to do in the absence of a determinate ideal of the person? To escape relativism, she suggests,

> we must rely not merely on the abstract principles that could be accepted (or rejected) by any plurality of minimally rational agents of indeterminate mutual independence. We must rely on specific interpretations of these principles that can be accepted or rejected by those actually involved. Only action and policies that guarantee the refusability by actual agents of offers and involvements that others propose can ensure consent that is not merely nominal but legitimating." (257)

We must appeal to *possible consent*, rather than either hypothetical or actual consent. We need to be sure that even those who are ignorant or weak could have "refused or renegotiated the roles or tasks that action on that principle imposed on them." (258) "Neither the apparent consent of the vulnerable nor the hypothetical consent of the glitteringly self-sufficient legitimates." (258) O'Neill contends that once we see this, we can also see that applying principles of justice will demand that we do more to be just to the vulnerable and ensure legitimating consent. "By contrast both idealized and relativized accounts of justice tend to conceal the fact that justice to the weak demands more than justice to the strong. Idealized accounts of justice tend to ignore actual vulnerabilities and relativized accounts to legitimate them." (258)

For reasons already given, I think that this criticism of idealized accounts of justice is mistaken. But what about O'Neill's appeal to possible consent? It isn't clear what is involved in ensuring possible consent. We might stand ready, for example, to alter our plans in the face of opposition from the vulnerable, but it might still be the case that they are unable to express opposition, for example, because of limitations of information, rationality, or cognitive functioning. Can we make it possible for them to consent short of removing the vulnerabilities that are the basis of our obligations? And what about

the bulk of cases in which this is simply not within our power? I confess that I find it hard to see how possible consent can be legitimating in the face of genuine human vulnerability. And it is this legitimation, I fear, that O'Neill foregoes by avoiding appeals to an ideal of the person. Surely what we want to know if we are concerned about legitimating consent is not simply whether it is possible for the vulnerable to consent, but what they *would* consent to or resist if they were able to see their situation aright. But this returns us to the kind of hypothetical question that calls for idealization.

IV. CONCLUSION

As I see it, there is not much hope in any case of establishing constructivism as a position distinct from either realism or relativism, so O'Neill's apparent failure to do so is ultimately not important.[12] But then I don't see this as the problem for constructivism that Rawls' critics evidently did and that many still do. If there is a takeaway from the other problems I have suggested for her Kantian Constructivism, it is that we cannot have sufficient determinacy without idealization and we cannot have legitimacy—or as we now say it, normativity—without sufficient determinacy.

12. My own view is that constructivism is a form of realism but that this in no way diminishes its interest.

REFERENCES

O'Neill, O. (1989), 'Constructivisms in Ethics', Presidential Address, *Proceedings of the Aristotelian Society* 89: 1-17.

Pettit, P. and Smith, M. (1993), 'Brandt on Self-Control' in B. Hooker, ed., *Rationality, Rules, and Utility: New Essays on the Moral Philosophy of Richard B. Brandt*. Boulder, CO: Westview. pp. 33-50.

Railton, P. (1986a), 'Moral Realism', *Philosophical Review* 95: 163-207.

_____ (1986b), 'Facts and Values', *Philosophical Topics* 14: 5-31.

Rawls, J. (1971), *A Theory of Justice*. Cambridge: Harvard University Press.

_____ (1980), 'Kantian Constructivism in Ethics', *The Journal of Philosophy* 77: 515-572.

_____ (1985), 'Justice as Fairness: Political, Not Metaphysical', *Philosophy & Public Affairs* 14: 223-151.

Shafer-Landau, R. (2003), Moral Realism. Oxford: Oxford University Press.

Street, S. (2010), 'What is Constructivism in Ethics and Metaethics?', *Philosophy Compass* 5/5: 363–384, DOI 10.1111/j.1747-9991.2009.00280.x.

THE RELATION BETWEEN MORAL THEORY AND METAPHYSICS

by Robert Stern

Robert Stern is a Professor in the Department of Philosophy at the University of Sheffield, UK. His publications include *Hegel and the Phenomenology of Spirit* (2nd edn Routledge, 2013), *Hegelian Metaphysics* (OUP, 2009), *Understanding Moral Obligation* (CUP, 2012) and *Kantian Ethics* (OUP, 2015). He is currently working on a monograph on the Danish philosopher and theologian K. E. Løgstrup.

by Robert Stern

THE CENTRAL QUESTION with which I shall be concerned in this paper is: how much can metaphysics contribute to moral philosophy? On the one hand, Derek Parfit has claimed that it can contribute a great deal; on the other hand John Rawls has claimed that it can contribute very little. In getting clear about the background to the dispute, and the contrasting viewpoints on which these respective positions are based, I hope to show that there are good grounds for agreeing with Parfit on this matter, and rejecting Rawls' claim that moral philosophy is independent of metaphysics.

However, it is necessary at the outset to be more specific about the nature of Rawls' claim, in order to be clear about the exact grounds of disagreement between himself and Parfit. In the first place, in arguing that metaphysics can contribute little to moral philosophy, I do not think Rawls should be read as denying that metaphysical issues can have a bearing on what might be called our moral *practice*, on the way we decide to treat other people, the natural world, and so on. Thus, to take an obvious example, I assume that Rawls would not deny that if it could be established that free will was illusory, this would affect our practices of punishment and reward, and alter our moral attitudes.[1] Secondly, I take it that Rawls is also not denying that some moral questions might perhaps be settleable by metaphysical arguments. So, for example, if there were some convincing metaphysical grounds for drawing a distinction between foetuses and persons, this could resolve the matter regarding the morality of abortion. Thirdly, I also assume that Rawls does not doubt that the development of our moral views, including our conception of the good and the right, has as its historical background an accompanying development in our metaphysical outlook, which Charles Taylor has recently called our 'moral ontology'.[2] Setting these interpretations of Rawls' position aside, therefore, when he suggests that metaphysics can contribute little to moral philosophy his claim is more limited than it might at first appear: using Rawls' own terminology, it is merely this, that metaphysics cannot contribute to moral *theory*, where by the latter he means the systematic comparison of competing moral conceptions, of which Kantianism and utilitarianism are the best known.[3]

1. For a discussion on this issue see P.F. Strawson, 'Freedom and Resentment', *Proceedings of the British Academy*, Vol. xiviii, 1962, pp.1-25, and (more recently) Ted Honderich, *The Consequences of Determinism: A Theory of Determinism*, Volume 2 (Oxford: Oxford University Press, 1990).

2. Charles Taylor, *Sources of the Self* (Cambridge: Cambridge University Press, 1989), pp.5-11. Cf. Hilary Putnam's observation that 'we have had to revise our theory of the good (such as it is) again and again as our knowledge has increased and our world-view has changed' (*Reason, Truth and History* (Cambridge: Cambridge University Press, 1981), p.215.

3. See John Rawls, "The Independence of Moral Theory', *Proceedings and Addresses of the American Philosophical Association*, Vol. 48, 1974-5, pp.5-22, pp.9-10.

Having narrowed the focus of our original question somewhat, I now wish to examine the grounds for the dispute between Parfit and Rawls over the relation between moral theory and metaphysics.

I.

Parfit's position arises in general from his 'hope' (as he expresses it) that '[much] of what is bad [in our emotions and attitudes] depends upon false belief'.[4] More substantially and more specifically, Parfit argues (in his book *Reasons and Persons*) that '... if we change our view about the nature of personal identity, this may alter our beliefs both about what is rational, and about what is morally right or wrong'.[5] Parfit's claim is that as regards the nature of persons, 'the truth here is very different from what most of us believe',[6] and that once the correct picture is acknowledged (i.e. his own), then our moral views must be changed, and changed for the better. Parfit's revisionist thesis regarding the nature of persons is well known: he rejects what he calls the Non-Reductionist View, according to which identity is 'all or nothing' and 'a person is a separately existing entity, distinct from his brain and body, and his experiences', and argues instead for what he calls the Reductionist View, according to which the identity of a person is just a question of physical and psychological continuity and connectedness, and a person is not a separately existing entity, over and above 'the existence of his brain and body, and the thinking of his thoughts, and the doing of his deeds, and the occurrence of many other physical and mental events'.[7] I do not propose to discuss the correctness or incorrectness of this metaphysical claim here. What interests me is rather the way in which Parfit then goes on to argue for certain radical ethical claims on the basis of his metaphysical thesis regarding the nature of persons, as this is the move of which Rawls is critical.

In fact, Parfit addresses several ethical issues (including questions of paternalism and desert); but given his conception of moral theory as outlined above, it is only with Parfit's treatment of utilitarianism that Rawls is concerned. As is well known, Parfit uses his metaphysical thesis regarding the person to defend utilitarianism against the charge (made by Rawls and others) that in seeking merely to maximise the good, it is insensitive to the *distribution* of benefits and burdens, and may allot great disutility to some, in order to achieve maximization overall. In his defence of utilitarianism, Parfit argues that if we accept his view that 'a person's life is less deeply integrated than most of us assume',[8] then 'we may believe that, when we are trying to relieve suffering, neither persons nor lives are the morally significant units',[9] but solely the states people are in at particular

4. Derek Parfit, 'Personal Identity', *Philosophical Review*, Vol. LXXX, 1971, pp.3-27, p.27.

5. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), p.306.

6. Ibid., p.450.

7. Ibid., p.275.

8. Ibid., p.336.

9. Ibid., p.341.

times. From this, two consequences follow: first, it no longer seems possible to compensate for past sufferings over a life, and secondly, 'it becomes more plausible to be more concerned about the quality of experiences, and less concerned about whose experiences they are'.[10] As a result, Parfit suggests, 'we should give less weight to distributive principles',[11] particularly the Principle of Equality, and that 'when we cease to believe that persons are separately existing entities, the Utilitarian view becomes more plausible'.[12]

Now, as I have said, I do not propose to discuss the correctness of Parfit's views regarding the nature of the person. Rather, my reason for outlining his position is merely to highlight the way in which he uses a particular metaphysical thesis to argue for a certain moral outlook, and to claim that common ethical beliefs that many of us hold are wrong, in so far as they are based on a false conception of the self. For Parfit, clearly, metaphysics has a central place in moral theory.

II.

However, in the work of John Rawls one finds a whole-hearted rejection of the outlook adopted by Parfit. In his article 'The Independence of Moral Theory', Rawls explicitly takes issue with Parfit over the importance of metaphysical questions concerning the nature of persons for ethics, supporting his negative claim by giving three specific criticisms and one general methodological argument.

I will begin by discussing the specific criticisms, which are as follows. First, while all moral conceptions, including Kantianism and utilitarianism, must accept those criteria of identity which are laid down by the philosophy of mind, 'the conclusions of the philosophy of mind regarding the question of personal identity do not provide grounds for accepting one of the leading moral conceptions rather than another'.[13] Second, while all moral conceptions use different ideals of the person (in the sense that they 'regard persons differently and prize different aspects of their nature'[14]), choice of an ideal is not determined by metaphysical considerations. Thirdly, the reasonableness of a moral conception depends on the kind of person it encourages us to become and the sort of society it leads us to create, and 'on neither of these questions is the problem of personal identity, as a problem in the philosophy of mind, likely to have much to say'.[15] Putting all these three claims together, Rawls' argument may be stated as follows: although the philosophy of mind may have things to say regarding the metaphysical question of personal identity, such questions leave our choice of moral conception undeter-

10. Ibid., p.346.

11. Ibid., p.346.

12. Ibid., p.342.

13. Rawls, "The Independence of Moral Theory', p.15.

14. Ibid., p. 17.

15. Ibid., p.15.

mined; while on ideals of the person and of society, on which moral conceptions really do differ, the philosophy of mind is silent, with the result that it can contribute nothing to moral theory.

As David Brink has pointed out,[16] Rawls' argument here depends on a distinction within theories of the person between accounts of personal identity and accounts of ideals of the person, for while the former are metaphysical in nature and belong to the philosophy of mind, the latter belong within ethics and political philosophy, and so are not independent of moral theory. Rawls suggests that the crucial difference between Kantianism and utilitarianism (in particular) lies in the ideal of the person they employ, and it is on this issue, and not on issues surrounding personal identity, that the choice between them lies. Utilitarianism, Rawls argues, operates with an ideal that treats persons as 'passive carriers of desires',[17] who live in a society whereby all social co-operation is co-ordinated to realize the end of maximizing want-satisfaction, whereas on the Kantian picture, persons are viewed as actively engaged in divergent plans of life, so that 'citizens of such a society pursue many different and opposed final ends',[18] with the result that society cannot be organized around any single rational good, such as want-satisfaction, but only around a shared conception of justice. The utilitarian ideal of the person means that 'all their conceptions of the good are publicly commensurable via a shared highest-order preference as to what is desirable; and so in this important respect the distinctiveness of persons is lost',[19] while the Kantian treats the individual as a rational, autonomous agent who does not want their freely chosen ends to be subordinated to some general social good, for this would be to lose their integrity as persons.

Now, Rawls claims that the moral conception we should abide by is dependent on 'the kind of person we might want to be and the form of society we might wish to live in and to shape our interests and character',[20] so that the choice between utilitarianism and Kantianism is determined by how we view the respective ideals of the person. Rawls clearly believes that none of us would wish to be expected 'to consider any new convictions and aims, and even to abandon attachments and loyalties, when doing this promises a life with greater overall satisfaction, or well-being, as specified by a public ranking',[21] and we should therefore reject the utilitarian ideal of 'bare persons',[22] and with it the associated utilitarian moral conception. Instead, Rawls suggests, we should accept the Kantian ideal of 'autonomous persons who have certain fundamental interests that they seek to advance',[23] and who simply wish to live in a society in which these divergent ends can be socially reconciled

16. David O. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), p.313.

17. John Rawls, 'Social Unity and Primary Goods', in *Utilitarianism and Beyond* ed. Amartya K. Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp.l59-85, p.l69.

18. Ibid., p.l82.

19. Ibid., p. 180.

20. John Rawls, 'A Weil-Ordered Society', in *Philosophy, Politics and Society*, Fifth Series, ed. Peter Laslett and James Fishkin (Oxford: Basil Blackwell, 1979), pp.6-20, p.6.

21. Rawls, 'Social Unity and Primary Goods', p. 181.
22. Ibid., p. 180.

23. Rawls. 'The Independence of Moral Theory', p. 18.

and fairly regulated. Rawls' claim therefore seems to be this: if we prize our freedom to pursue our own goals and do not wish to live in a society that treats individuals as inter-substitutable by reducing them to carriers of wants and desires, we should adopt the Kantian approach, and abandon utilitarianism as a moral conception.

Assuming, then, that this is the proper way to understand what Rawls himself admits are his rather 'allusive' observations in 'The Independence of Moral Theory', we may now inquire into the plausibility of his position. I will argue that Rawls wrongly locates the grounds on which we should decide between Kantianism and utilitarianism as moral conceptions, and that how one makes the real ethical choice does depend on how one decides between two competing metaphysical pictures of the self.

Rawls, as we have seen, tries to argue against utilitarianism as a moral conception on the grounds that it adopts an unacceptable ideal of the person, one which asks us to subordinate all our individual aims to the overall social goal of maximizing utility, and allow ourselves to be compared against everyone else merely in terms of our capacities for satisfaction. His claim is that this conception incorporates a simplified and homogeneous view of what it is to be a person, while Kantianism has greater room for the value we place on our individual projects and attachments, and keeps the burden of moral commitments within reasonable bounds, by not expecting individuals 'to acquiesce in a loss of freedom over the course of their life for the sake of a greater good enjoyed by others'.[24] Rawls therefore makes the contrast between these ideals the primary grounds on which we are supposed to choose between Kantianism and utilitarianism.

Is this plausible, however? It assumes that our selection of a moral conception is primarily dependent on the kind of person it expects us to become, coupled with the claim that not many of us are likely to want to be treated as the sorts of moral agent envisaged within utilitarianism. Now of course, from the writings of Bernard Williams and others we are familiar with the claim that utilitarianism is overstrenuous, ignores the value of integrity, and does not take seriously enough the place in our lives of personal relations and projects, and if Kantianism were any better in this respect it might be a factor in its favour; but, as Bernard Williams has also argued, this seems unlikely.[25] More importantly, it is not clear that in comparing the competing moral conceptions of Kantianism and utilitarianism, this is a difference that matters. What we are looking for here is (to use Sidgwick's famous phrase) a 'method of ethics',

24. John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1972), p. 176.

25. See Bernard Williams, 'Persons, Character and Morality', in his *Moral Luck* (Cambridge: Cambridge University Press, 1981), pp.1—19, and *Ethics and the Limits of Philosophy* (London: Fontana, 1985), Chapter 5.

that is 'any rational procedure by which we determine what individual human beings "ought"—or what it is "right" for them—to do, or seek to realise by voluntary action',[26] and it is on this issue that utilitarianism and Kantianism should really be compared, not on their respective ideals of the person. Rawls' emphasis on the latter in deciding on a moral conception is therefore misplaced: rather than trying to decide which embodies the more plausible ideal of the person, what in fact needs to be shown is that one of them offers a better account of what we ought to do, of what is right and wrong, and of what is of moral value.

Now, of course, on this question utilitarianism and Kantianism differ too, and in particular many have argued that utilitarianism is an inadequate moral conception in so far as it places no weight on the demands of equality and fairness, and only gives value to maximization. However, I will now suggest that (contra Rawls) in appealing to the ethical notion of equality the Kantian relies on a metaphysical picture of the person, which itself also derives from Kant, and contrasts with the picture adopted by the utilitarian.

The principle of equality gains its plausibility from the claim that in deciding how to distribute goods we should take into account how that distribution will affect each individual person, and not just how it will affect the level of utility taken over society as a whole. The metaphysical account of the person that underpins the principle of equality is fundamentally *perspectival*, in so far as it emphasises that what it is to be a person is to have a certain view on the world, to be a centre of experience, and it is from this that the separateness of each self is said to arise. This implies that the individual 'is owed an effort at identification: that he should not be regarded as the surface to which a certain label can be applied, but one should try to see the world (including the label) from his point of view'.[27] It is this metaphysical account of the self that underlies the Kantian claim that 'imaginatively one must split into all the people in the world, rather than turn oneself into a conglomeration of them';[28] and in being required to take each person's viewpoint separately, the person (*contra* Parfit) becomes a morally significant unit on which the Kantian can base a claim for equality.

Against this, Parfit argues that while we may be able to distinguish between persons at a time, once we see that taken as a whole a person's life lacks any deep unity, we should come to see this distinction as less morally significant. As a consequence, he is happy to place more emphasis on the overall quality of experiences, and less on their distribution, claiming that his Reductionist ac-

26. Henry Sidgwick, The Methods of Ethics, 7th edn, reprinted (Indianapolis: Hackett, 1981), p.1.

27. Bernard Williams, 'The Idea of Equality', in his *Problems of the Self* (Cambridge: Cambridge University Press, 1973), pp.230-49, p.236.

28. Thomas Nagel, 'Equality', in his *Mortal Questions* (Cambridge: Cambridge University Press, 1979), pp. 106-27, p. 127.

count of the person thereby lends support to the utilitarian view, not by conflating persons, but as a result of their partial disintegration.[29]

Now, what matters here is not the validity or otherwise of Parfit's metaphysical account of the person (although I think he does not properly refute the Kantian position outlined above, which should be distinguished from the Cartesian approach he mostly considers). More importantly, my purpose has been to suggest that how we decide between Kantianism and utilitarianism concerns the extent to which we take seriously the moral claims of the person, and that how we judge these claims is crucially dependent on what we take the nature of the person to be.[30] If I am right, none of Rawls' arguments considered so far establish the independence of moral theory from metaphysics; his general, methodological criticism of Parfit's approach now remains to be considered.

III.

In essence, Rawls' general, methodological objection to Parfit's approach is that it violates the method of reflective equilibrium on which moral theory is based, and by which progress in moral philosophy is made. According to the method of reflective equilibrium, the only way of deciding between competing moral conceptions (like Kantianism and utilitarianism) is by testing them against our considered moral judgements, which constitute a 'definite if limited class of facts against which conjectured principles can be checked'.[31] Parfit goes too far, however, in using metaphysical arguments to challenge these considered moral judgements, and as such adopts the kind of revisionist approach that is at odds with the essentially descriptive method of reflective equilibrium. This important claim deserves to be considered in more detail.

The technique of reflective equilibrium (which has its roots in Rawls' reading of Sidgwick[32]) is a method which uses only ethical considerations to decide between rival moral conceptions. Beginning from the thought that we all have some views on what is morally or politically right and wrong, just and unjust, acceptable and unacceptable, and that we are capable of judging the conviction with which these views should be held, Rawls argues that these moral judgements provide a background against which moral conceptions can be tested. Any successful moral conception must be capable of fitting, extending and explaining our considered ethical views, and although the latter are not immune from revision entirely, the assumption of the method is that no moral conception fundamentally at odds with our considered

29. Parfit, *Reasons and Persons*, p.336.

30. It should perhaps be emphasised that Parfit's central concern is with the nature of persons, not just their identity: he simply uses considerations about personal identity to support his Reductionist View, by arguing that as personal identity over time just involves physical and psychological continuity, persons cannot be substantial egos. One reason why Rawls underestimates the interest of Parfit's position may be that he mistakenly takes Parfit's primary claim to be about the identity, and not the nature, of persons.

31. Rawls, *A Theory of Justice*, p.51.

32. See ibid., p.51 note, and 'The Independence of Moral Theory', pp.9-10. Rawls' interpretation of Sidgwick is influenced by J.B. Schneewind, 'First Principles and Common Sense Morality in Sidgwick's Ethics', *Archiv für Geschichte der Philosophie*, Vol. 45, 1963, pp. 137-56. For a critique of Rawls' interpretation see Peter Singer, 'Sidgwick and Reflective Equilibrium', *The Monist*, Vol. 58, 1974, pp.490-517.

moral judgements should be taken seriously. So, for example, a good reason for ruling out utilitarianism is that 'the principle of utility is incapable of explaining the fact that in a just society the liberties of equal citizenship are taken for granted, and the rights secured by justice are not subject to political bargaining nor to the calculus of social interests';[33] by contrast, the Kantian conception of fairness is able to 'express the principles of justice which stand in the background and control the weights expressed in our everyday judgements'.[34] It is on this need to achieve a coherent fit between our moral judgements and our chosen moral conception that Rawls' method of reflective equilibrium relies.

Now, Rawls' general methodological objection to Parfit is that the latter fails to adopt this procedure of testing our moral conceptions to see which provides us with the most internally coherent moral scheme; instead, Parfit begins from outside morality altogether, and tries to get us to revise our considered moral judgements in the light of his metaphysical claims. Thus, while Rawls does accept that our moral views can be altered in the light of our search for a systematic moral outlook,[35] his method does not allow for the much stronger Parfitean claim that 'if we change our view about the nature of personal identity, this may alter our beliefs ... about what is morally right or wrong'.[36] To do so, Rawls would have to accept that it may not be enough to test moral conceptions purely within 'the structure of our moral sensibility',[37] but that they may also need to be matched against our best metaphysical beliefs about the world, with the implication that metaphysics can no longer be set outside moral theory.

It would seem, then, that Rawls' general methodological argument that moral theory is independent of metaphysics simply begs the question against Parfit: for it assumes that the issue between utilitarianism and Kantianism can be settled merely by going back and forth between our moral judgements and conceptions, whereas it is Parfit's claim that it is precisely these judgements that in the light of metaphysical considerations have to be changed. Rawls' apparent assumption that moral judgements must remain immune from changes in our metaphysical beliefs just seems to rest on the unmotivated and implausible claim, that our moral reflection forms (so to speak) a 'closed circuit', encompassing moral judgements and moral conceptions alone.[38]

However, it would be unfair to Rawls to suggest that this picture of reflective equilibrium represents his actual procedure: for in *A Theory of Justice* our moral judgements and moral conceptions are not just tested against each other, but are also set against a third varia-

33. John Rawls, 'Distributive Justice', in *Philosophy, Politics and Society*, Third Series, ed. Peter Laslett and W.G. Runciman (Oxford: Basil Blackwell, 1969), pp.58-82, p.59.

34. Ibid., p.81. See also *A Theory of Justice*, pp.49-50.

35. See for example, Rawls, "The Independence of Moral Theory', p.8.

36. Parfit, *Reasons and Persons*, p.306.

37. Rawls, "The Independence of Moral Theory', p.7.

38. For an attempt to introduce metaphysical principles, as well as moral judgements and conceptions, into the procedure of reflective equilibrium, see Vinit Haksar, *Equality, Liberty and Perfectionism* (Oxford: Oxford University Press, 1979).

ble, namely the contractual situation, and Rawls clearly sees a close interrelation between all three. Thus, our conception of reasonable contractual conditions may be influenced by our considered moral judgements, which both may determine our choice of moral conception; but equally the balance may go the other way, until some overall stability is reached. It is important to note, therefore, that Rawls' argument against utilitarianism is not just that it does not fit with our considered moral judgements: it would also not be reasonable to expect it to be chosen as a principle of justice by parties in the contractual situation of the original position.[39]

Now, once this contractual element is introduced, this might seem to give Rawls' argument against Parfit a much higher degree of plausibility: for, in deciding what principles of justice might be chosen in the contractual situation, metaphysical beliefs about the nature of the person might seem unlikely to influence the choice. The aim, after all, is simply to arrive at principles that (as Scanlon puts it) 'no one could reasonably reject as a basis for informed, unforced general agreement';[40] and metaphysical considerations regarding the nature of the person would seem to have no bearing on Rawls' fundamental claim against utilitarianism, that 'a rational man would not accept a basic structure merely because it maximized the algebraic sum of advantages irrespective of its permanent effects on his own basic rights and interests',[41] though he would accept a modified notion of equality (the Difference Principle). Thus, moral theory would seem to be independent of metaphysics, for the latter could be said to have no effect on what principles of justice we could reasonably expect others to accept or choose in the contractual situation.

However, Parfit argues that metaphysical issues regarding the nature of the person would and should have a *direct* bearing on this question: for if we came to abandon the Non-Reductionist view of the person, we *should* rationally come to accept that whether I or someone else benefits does not matter, but only the extent of the overall gain. We should also come to accept other revisions in what we would accept, such as greater paternalistic intervention and punishment for imprudence, on the basis of Parfit's view of the self. The problem for Rawls here is this: Parfit does seem justified in suggesting that what principles of justice we might agree to, and what principle we could reasonably insist other contractors should accept, is dependent on how we view the nature of persons, and not just on the state of our information in the original position. In particular, as Parfit himself argues, the extent to which we are prepared to accept self-interest as a reasonable defence against the sacrifices required for the sake of general welfare, and the extent

39. See Rawls, *A Theory of Justice*, pp. 183-92.

40. T.M. Scanlon, 'Contractualism and Utilitarianism', in *Utilitarianism and Beyond*, ed. Amartya K. Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp.103-28, p.110.

41. Rawls, *A Theory of Justice*, p. 14.

to which we as individuals pursue the former in favour of the latter, will depend upon how we view the unity of ourselves as persons and the unities and differences within and between lives. Rawls' contractual argument relies on the thought that as none of us is a perfect altruist it is not reasonable to expect us to accept utilitarianism in the original position; but it is precisely such an altruistic outlook that Parfit claims his metaphysical arguments will lead us to adopt, and we will then come to see the reasonableness of utilitarianism as an ethical position.

IV.

Finally, something must be said about the *motivation* for Rawls' claim, that moral theory is independent of metaphysics.

Clearly, Rawls' central aim is to abandon any reliance on metaphysics as 'first philosophy', and therefore he denies that 'ethics awaits an answer to such problems as those of the freedom of the will and personal identity'.[42] In this respect Rawls' position may be compared to Sidgwick's, who was also exercised by 'the very important question whether ethical science can be constructed on an independent basis',[43] and tried to show that questions concerning free will are irrelevant to deciding between the methods of ethics. Moreover, just as for Sidgwick 'a policy of keeping non-ethical commitments of moral philosophy to an absolute minimum…[arose] from the methodological principle of aiming at general agreement',[44] so Rawls clearly holds that only when ethics and politics are free of highly controversial yet (perhaps) unsettleable claims regarding metaphysics and the nature of persons will a stable, democratic, pluralistic consensus be possible.[45]

This might reasonably be questioned, however. After all, one plausible explanation of the increasingly liberal outlook of modern cultures is our convergence on one conception of the person, and it is this convergence that underpins and makes possible the consensus. Rawls suggests that a public basis for agreement cannot be supported by any metaphysical claims: but, it might be argued, there is in fact a *greater* chance of agreement based here, than on our considered moral judgements alone, on which (as we have seen) Rawls' approach depends. Moreover, while Rawls may be right to deny that metaphysics should always be treated as methodologically prior to ethics, how one settles the dispute between Kantianism and utilitarianism can be seen to rest on metaphysical issues, if it is indeed true (as Rawls himself suggests)[46] that the question turns on how one views the separateness of persons.

42. Rawls, 'The Independence of Moral Theory', p.6.

43. Sidgwick, *The Methods of Ethics*, p.507. For a discussion of this issue in Sidgwick, see J.B. Schneewind, *Sidgwick's Ethics and Victorian Moral Philosophy* (Oxford: Oxford University Press, 1977), pp.204-14.

44. Schneewind, *Sidgwick's Ethics*, p.204.

45. See for example John Rawls, 'Justice as Fairness: Political not Metaphysical', *Philosophy and Public Affairs*, Vol.14, 1985, pp.223-51.

46. Rawls, *A Theory of Justice*, p.27.

A further important motivation for Rawls' scepticism regarding metaphysics (which is also perhaps derived from Sidgwick) concerns Parfit's claims regarding moral progress. Parfit is of the view that a good way of making headway in ethics is by arriving at a more accurate metaphysical picture of ourselves and the world, which may then lead us to adopt a better moral conception. For Rawls, by contrast, 'the further advance of moral philosophy depends upon a deeper understanding of the structure of moral conceptions and of their connections with human sensibility',[47] and progress can only be achieved using the 'internalist' method of reflective equilibrium, of working at the development of a more coherent fit between our moral judgements and our moral conceptions. On Rawls' view, therefore, the evolution of our moral conceptions is not dependent on the development of our metaphysical understanding of the world, but is rather brought about through the development of greater coherence within our moral outlook.

Now, in suggesting that deficiencies in our moral views can only be overcome through progress in our metaphysics; Parfit is making a claim that many, not just Rawls, would want to reject.[48] It would certainly constitute the basis for a very strong claim regarding the dependence of moral theory on metaphysics, for it would imply that no moral progress is possible unless it is supported by a better understanding of how things are. Ultimately, however, I do not believe that Parfit's emphasis on metaphysics rests on a dogmatic foundationalism and a misguided search for a 'first philosophy': rather, it simply stems from his 'hope' (referred to earlier) that '[much] of what is bad depends upon false belief: Parfit's suggestion is that if only we could arrive at a truer metaphysical picture of the world, we would be forced to revise a number of our ethical beliefs in such a way that we would then have a better understanding of what is right and wrong. It is therefore perhaps only if we can arrive at some satisfactory assessment of the grounds for this hope, that we can finally decide on the real contribution metaphysics can make to moral theory.

One minimal way of interpreting Parfit's position is this: that deficiencies in moral outlook have often been based on mistaken metaphysical beliefs, and that if one reflects on what occurs when cultures or individuals undergo changes in ethical outlook, it is clear that accompanying shifts in their metaphysical views are of the utmost significance. Once such shifts have occurred, a previous moral outlook may no longer be a 'real option',[49] and may bring about a complete alteration in our views of what is right and wrong. In this way, it is possible to claim that metaphysics constitutes a very important part of our moral framework, for when our

47. Rawls, 'The Independence of Moral Theory', p.22.

48. Cf. Richard Rorty's remark that '... intellectual and moral progress (is) a history of increasingly useful metaphors rather than the increasing understanding of how things really are'. (*Contingency, Irony and Solidarity* (Cambridge: Cambridge University Press, 1989), p.9.)

49. Bernard Williams' phrase: see 'The Truth in Relativism', in his *Moral Luck* (Cambridge: Cambridge University Press, 1981), pp. 132-43.

metaphysics changes it may leave parts of our moral life exposed, leading us to adopt a new moral position, and to see our previous one as flawed.

An obvious example of such a phenomenon is the change in our ethical attitude to the animal world. Whereas the cruel and exploitative treatment of animals had a stable metaphysical and theological foundation up until the eighteenth century, the remarkable enlargement of moral concern to include other species that occurred in England from the 1740s onwards can only be explained by the new developments in metaphysical and theological outlook that took place from the same time, and which included the acknowledgement that animals feel pain and are not merely machines, that the world may not have been created for the benefit of men alone, and that there may be no metaphysical distinction between animals and ourselves which is capable of justifying our treatment of them.[50] In this case, it could be argued, a new (and better) understanding of what is right and wrong as regards the moral claims of animals was made possible and can be explained by a new (and better) metaphysical understanding of how things are.

Now, of course, this is not to deny that moral progress could be brought about and explained in other ways, for example by appeal to changing social circumstances, or to 'wider experience, fuller knowledge, more extended and refined sympathies'.[51] The claim merely is this: that if the aim of moral philosophy is ultimately to see how we can make moral progress, to improve our understanding of what is right and wrong, one plausible route may be to examine the metaphysical beliefs that underpin our moral reasoning; for if false metaphysical beliefs have been responsible for deficiencies in the moralities of the past, it is reasonable to assume that they may be responsible for deficiencies in the morality of the present. In this way, as Parfit himself suggests, metaphysical inquiry may contribute a good deal to the revision of our moral outlook, and so (contra Rawls) may be said to form an indispensable part of moral philosophy.[52]

50. For a fascinating historical study of this change in moral outlook (on which I have drawn), see Keith Thomas, *Man and The Natural World: Changing Attitudes in England 1500–1800* (Harmondsworth: Penguin Books, 1984).

51. Henry Sidgwick, *Lectures on the Ethics of T.H. Green, H. Spencer, and J. Martineau* ed. by E.E. Constance Jones (London: Macmillan, 1902), p.352; cited Schneewind, *Sidgwick's Ethics*, p. 213.

FROM METAPHYSICS TO ETHICS VIA PHILOSOPHY OF MIND: THREE ROUTES

Commentary by Sophie Grace Chappell

SOPHIE GRACE CHAPPELL is Professor of Philosophy at the Open University. She has previously taught at universities including UEA, UBC, Manchester, and Oxford. Her books include *Reading Plato's Theaetetus*, *Ethics and Experience*, and *Knowing What To Do*. Her current research engages with the idea of a philosophical ethics based in experience rather than on moral theory, and in particular in epiphanic experiences.

FROM METAPHYSICS TO ETHICS VIA PHILOSOPHY OF MIND: THREE ROUTES

Commentary by
Sophie Grace Chappell

HOW DOES OUR METAPHYSICS AFFECT OUR ETHICS?[1] How should it? As Robert Stern notes in 'Metaphysics and Ethics' (1992) John Rawls and Derek Parfit apply these questions specifically to the metaphysics of personal identity. Very early in the unfolding of his theory Rawls complains—in *A Theory of Justice* the complaint is decidedly epigrammatic, and not much defended or commented on—that "Utilitarianism does not take seriously the distinction between persons" (Rawls 1971: 26-7).

This complaint is the *point de départ* for Parfit's reductionism about persons. As Buddhist ethics is largely about dispelling *maya*, the illusions (as Buddhists take them to be) of fixity that chain us to desire, fear, grief, and other attachments; so Parfit, who himself notes the parallel with Buddhism (1984: appendix J), argues over hundreds of pages that a true metaphysics of persons will centrally include what he calls "R-relations", relations of psychological connectedness and continuity. It will *not*, centrally or underivatively, include *persons*. For there are "no further facts" about personhood that are not already stated, as fully as they intelligibly can be, by stating all the facts about R-relatedness. Moreover, Parfit presents thought-experiments designed to show that R-relatedness can hold in ways that compromise or downright subvert the intuitive metaphysics of personal identity that Rawls apparently takes for granted.

So for Parfit, Rawls' "separateness of persons" is simply an illusion. Personhood is not real (metaphysical premiss); so personhood cannot be "what matters" (moral conclusion). What *does* matter is the states of R-related pain or pleasure, happiness or unhappiness, that we find ourselves in at particular times. To these states the morally-required observer-response is compassion, empathy; note again the striking parallel with Buddhism's notions of *karuńā* and *mettā*. In the history of European moral theory, compassion and empathy, Parfit evidently thinks, are most clearly the keynotes of the utilitarian tradition; this is partly why he places himself within that tradition.

So goes Parfit's anti-Rawlsian argument from metaphysics to ethics. One response is to query the sense of his talk of "states of R-related pain or pleasure" as

separable time-slices of experience. Experience seems essentially diachronic: it seems to be constitutively shaped by what Husserl might call its relations of *protention* backwards and forwards in time. But these protentions are relations whose full nature and significance Parfit's talk of "R-relations" does not even look like capturing. This is a line I have developed myself against Parfit (e.g. in Chappell 1998a chapter 4, and in Chappell 2003).

Another response is this: "*If* thought-experiments can be constructed in which personhood-ascriptions become indeterminate in truth-value; and *if* the possibility of such thought-experiments is reason to conclude that there is something unreal about personhood-ascriptions; *then* it seems equally possible to construct thought-experiments in which R-relatedness-ascriptions become indeterminate in truth-value *too*." This line also is one I have argued, in a contribution of my own to the Proceedings of the Aristotelian Society (Chappell 1998b, and see also Chappell 1995). The line perhaps has the corollary that persons and R-relata are equally real; or as some Buddhists will say, equally unreal.

Rawls himself takes neither of these lines. What Rawls does say in response to Parfit has often—as Stern observes—had the air of simply begging the question, by insisting that metaphysical issues like those that Parfit raises *just cannot* affect ethics. It is an interesting question why Rawls seems confident enough in his own approach to ethics just to shrug off Parfit's counter-arguments. The reason, I think, is that Rawls is sure—or was sure when he wrote *A Theory of Justice*—that his own account of persons as rational bargainers behind a veil of ignorance (a) brings with it no heavyweight metaphysical assumptions and (b) is just the picture of persons that we as ethicists naturally want to work with.

This double assurance has come under heavy fire since 1971; and rightly so. The obvious problem is not that Rawls faces the same sort of troubles about the self as his father in philosophy Kant. Maybe at least some of the grave difficulties that attend Kant's fundamental commitment to the cloudy notion of a "noumenal self"[2] *do* descend to Rawls's account of persons. But if so, not obviously so; the point would take some arguing. The problem that *is* obvious is Rawls's failure—and this too is a distinctively Kantian failure—to philosophise historically about the self. *A Theory of Justice* presupposes a "thin theory" of the self, ahistorical and punctual,[3] as if rational bargaining were the truth about all selves anywhere and anywhen; as if there were *nothing more to say* about what selves essentially are. But any thin theory merits suspicion precisely because of its thinness; because of what it does *not* say.[4] What the conjuror di-

2. For unapologetically doughty argument that Kant is deeply committed on this front see Timmermann J. & Reath, A. 2010.

3. Charles Taylor's word: see Taylor 1989, e.g. p. 159.

4. Cp. my attack on the alleged thinness of "thin ethical properties" in Chappell 2014a.

rects our attention away from is just what we should look at most closely; when we engage with moral theorists nothing, as a rule, is less to be taken for granted than what the theorist invites us to take for granted. (So for instance Rawls's specific calculations of the choices of a rational agent in the original position depend upon a particular, and historically contingent, account of economic rationality; his theory's failures to escape from history run all the way down.) We should not miss just how deeply ideological Rawls's apparatus really is:— how much metaphysics it packs in simply by *silence and assumption*,[5] how non-accidentally perfectly that metaphysics of the person yields the result that the paradigm persons are able-bodied mature-but-not-old intelligent articulate well-educated self-assertive white heterosexual economically-independent male Westerners like John Rawls himself.

We have good reasons of moral—and political—mental hygiene to remind ourselves frequently of this familiar objection to Rawls. But "familiar" is the word. For anyone who like me is politically well to the left of Rawls, the objection is glaringly obvious; Rawls himself has been addressing it from *Political Obligation* onwards. Having got it off my chest (for now), I won't say more about it here.

Instead I'll take a less beaten path. I'll argue that while Rawls, Parfit, and Stern are right that the philosophy of personal identity is important for ethics, another part of our metaphysics is even more important, namely our account of mind. I give three items of evidence of its importance. We can label these (1) *animal minds*, (2) *imaginative identification*, and (3) *psychological concepts as ethically thick concepts*.

I. ANIMAL MINDS

Here is a patent truth: how we should treat animals depends on what animals are. One way to argue for this truth, if argument is needed, is to note that no serious debater of animal ethics ever disputes it. The debates are either about what animals are, or they are about *exactly* what duties follow from what various animals are like; no one disputes the *general* idea that animals' moral status is determined by animals' natures. Another way to argue for it is to point out that it's a substitution-instance of the universal truth that how we should treat *anything* depends on what it is; the claim about animals is simply a particular consequence of the general supervenience of ethical normativity on the facts. Presumably this general supervenience is the most global connection of all between metaphysics and ethics.

5. "The sound of a successful prejudice is silence"; Chappell 2010: 32.

When Stern briefly addresses the place of animals in metaphysics-to-ethics arguments, towards the end of his paper, perhaps he has Bentham's famous "Can they *suffer*?" (Bentham 1781, fn. 122 to chapter XVII) in mind:

> the remarkable enlargement of moral concern to include other species that occurred in England from the 1740s onwards can only be explained by the new developments in metaphysical and theological outlook that took place from the same time, and which included the acknowledgement that animals feel pain and are not merely machines. (287)

In fact I doubt that there was anything new, in England in 1740, about acknowledging that animals can suffer. There is evidence from long before then, and from societies much crueller than England,[6] that people in general have always been aware of this patent truth, and that no one except the odd philosopher has ever thought to deny it. Perhaps what changed or began to change in the eighteenth century was, rather, the sense that animal suffering at human hands was inevitable and therefore (?) tolerable. Or perhaps, more darkly, people at that time moved away from being prepared to *allow* themselves to enjoy cruelty, in the way that (according to Nietzsche) most pre-modern societies not only enjoyed cruelty but made a festival out of it: "No cruelty, no feast".[7]

Be that as it may, I suggest that "Can they suffer?" is not, anyway, the most ethically fruitful question about the metaphysics of animal life. No doubt Bentham is right that this is a better question than "Can they reason?" and "Can they talk?". But maybe a better question about animals than any of these is: "Are they viewpoints on the world?"

A viewpoint on the world is a subjectivity, a consciousness, a locus of experience, a "here", an "I", an index in the sense of "indexical". It is what I am; it is what you are; it is what anyone is, where "anyone" ranges over humans; it is also what many non-human animals are.

Reasons and Persons seems, incidentally, very unclear indeed about the relation between experiences and locuses of experience. It looks from pp.217 ff. as if Parfit thinks that there can be experiences "impersonally described", which apparently means experiences without locuses of experiences; it also looks—though the passage in question is extremely compressed and gnomic—as if Parfit thinks that these impersonal descriptions of experiences are the only really correct ones. But the idea of experiences without locuses of experience seems to me no more coherent than the idea of the angles of a triangle without the triangle itself. As far as I can see

6. One example is Pliny the Elder's story of Pompey and the elephants, *Natural History* 8.7, discussed in Chappell 2015.

7. Nietzsche (1887), essay 2, section 6. Compare Macaulay's famous observation, in Ch.11 of his *History of England,* that '*The puritan hated bear baiting, not* because it gave pain to the bear, but because it gave pleasure to the spectators.' (Macaulay 1849: chapter 11). Macaulay is not (as people often suggest) merely offering a quip or an epigram; he means this as a serious point about the puritans' uninterest in animal suffering. For the historical development of attitudes to animals in Britain, the whole passage repays study.

experiences are *necessarily* owned—like Cheshire-Cat smiles; or like properties in general, as indeed the word "property" suggests. Which, of course, is not in itself to say anything about the *nature* of their owner; nor is it to commit the elementary scope-fallacy that gets us from "For every experience, something is the owner of that experience" to "Something is the owner of every experience". This whole area seems very sketchily dealt with in *Reasons and Persons*, and indeed in utilitarian thinking about pleasure and pain and other experiences in general.

Plainly it is only viewpoints on the world that can suffer. But one reason why mine is a better morally-diagnostic question than Bentham's is because suffer (or not) is not the *only* thing that viewpoints on the world can do. The whole gamut of possible experience is open to any viewpoint on the world; we might say with only a bit of hyperbole that to be a viewpoint on the world is to be a *world*.[8] It is to be a version of the reality that is arranged around, or better *for*, that viewpoint; where "around" and "for" are loose and open, and potentially very inclusive indeed.

To put it as briefly as possible: to be a viewpoint on the world is to be a *you*. In these terms, the point about animals and animals' moral standing is that most of the animals that we most frequently and saliently encounter are, and are *obviously*, viewpoints on the world: *yous*. And it is because they are viewpoints on the world that they have moral standing. Their capacity for suffering is undoubtedly part of what it is for them to be viewpoints on the world; but only part. What does the key work in making animals morally significant (in just the way that they are) is their status as *yous*.

What happens when I empathise with another being, an animal for instance, is that I, a viewpoint on the world, become aware of another viewpoint on the world, and respond volitionally and affectively to that other viewpoint as such. With another human being this is typically mutual: when I become aware of my daughter Róisín as a viewpoint on the world, she simultaneously becomes aware of me, Sophie Grace, as a viewpoint on the world (and Griceanly, I become aware of her awareness, and she becomes aware of my awareness of her awareness, etc.). With most dogs it seems mutual too: it looks very much as if Jess the collie is aware of me as a viewpoint on the world when I am aware of her as one, and it may well be, also, that Jess is aware of my awareness of her as well. With simpler creatures, e.g. rabbits, it is harder to tell: when I interact with our blue minilop Vaisey I have no serious doubt that she is a viewpoint on the world, but I can't tell whether she realises that I am one too.

8. For this claim cp. Sprigge 1983.

There is more to say about this than I can say here; for now, let's just note two things. First, the metaphysical feature of being a viewpoint on the world is a crucial determinant of intuitive moral status. Secondly, therefore, any argument in metaphysics that has the upshot that there are not, and/or cannot be, any such things as viewpoints on the world will have serious ethical implications. Parfit's own view as sketched above may be one such view; perhaps any "no-ownership" view shares these implications. So, presumably, does any version of eliminative materialism that denies flat-out that there is, literally, "anything going on in the head": more about views like that under (III) below.

II. IMAGINATIVE IDENTIFICATION

Another crucial tool of intuitive moral thinking is *imaginative identification*, or as we informally put it, "putting myself in your shoes": the technique whereby I seek a better understanding of some moral situation involving you, by imagining that I *am* you in that situation. The Golden Rule found in the Gospels (Matthew 7.12) and elsewhere is an explicit codification of this moral technique.

I think our ordinary, intuitive moral thinking depends a great deal on exercises in imaginative identification, just as it does on the notion of a viewpoint on the world; I will suggest in a minute that the two ideas are intimately related. Given their importance in ordinary moral thought, it is surprising and puzzling how little either notion has been represented in most philosophical moral theory.

Part of the reason why must be the familiar point that both the two most important moral theories in the European tradition, utilitarianism and Kantianism, are in different ways *impersonal* theories of ethics.[9] Utilitarianism's fundamental suspicion of the notion of viewpoints on the world, and its familiar preference for what Bernard Williams calls "the absolute conception", comes out in many ways; one instance is Parfit's already-noted preference for an impersonal description, and a no-ownership theory, of experiences.

As for Kant, it is not clear that his moral theory gives any special place to viewpoints on the world as such. Perhaps that is partly what is meant by his talk of moral agents as "ends in themselves", or again by his talk of "dignity". But if so, Kant hardly spells this out; though of course he had successors like Fichte and Hegel who were ready to read just this sort of significance into what Kant does say.

Also, Kant is openly contemptuous of the "trivial

9. For classic argument to this effect see Williams 1981.

quod tibi non vis fieri" (*Groundwork*, Ak. 4:430, note). The Golden Rule, he says there, does not "contain the ground" of any duty: the fact that *you would not like it* "if someone did that to you" is never, or hardly ever, the reason why they should not do that to you. (If they shouldn't.)

This criticism of the Golden Rule, I suggest, almost entirely misses the mark. No doubt the Golden Rule *can* be treated in the way Kant condemns here, as a rule that simply enables us to compile and assess known and conjectured preferences. Perhaps it is treated that way, for instance, in Richard Hare's moral theory, or in the hybrid theory that Parfit ends up with in *On What Matters*, or even in Tim Scanlon's moral heuristic of reasonable rejection. But I want to suggest that there is a better use of the Golden Rule: an imaginatively deeper use, and a more morally serious one. I would like to think, but can't prove, that this use is what Jesus himself had in mind. Certainly something like it is sporadically visible elsewhere in the moral tradition that Jesus founded, for example in the ethically remarkable speech that Shakespeare famously gives to Shylock ("If you prick us, do we not bleed?": *Merchant of Venice*, Act 3 Sc.1).

On this reading of the Golden Rule, the focus is not on the preferences themselves, so much as the people behind them: the persons who *have* the preferences. The important thing is not that anyone happens to want or like or dislike this or that or the other. It is that dealing with other people means dealing with *beings who characteristically have* preferences, wants, likes, and so on, *just as I do myself*. So understood, the Golden Rule is not there to generate shopping-lists of wants for moral calculation to aggregate over. It is there to remind me of a morally crucial equality: that each and every person I encounter is a locus of desires and affections, fears and aversions, *in just the same sense as I am*. It is there to illuminate imaginatively for me what I ought, at least in the abstract, to know anyway: that to deal with other people is to deal with beings who are viewpoints on the world just like me. Which means dealing with beings just as poignantly fragile as me, and—in Rai Gaita's precisely-chosen word—just as precious (Gaita 2001: ch. 1).

As Stern reminds me, something like this conception of the Golden Rule as a spur to deeper and more compassionate imagining is there in the ethics of his recent main subject of study, the ethics of Knud Ejler Loegstrup, who has this to say in pretty well explicit reproof of Kant:

> The demand comes to expression in, for instance, the Golden Rule: Do unto others as you would have them do unto you. This is anything but a tepid rule of repaying, even if, taken literally, it might seem to be

such. On the contrary, it is a rule governing the use of the imagination. It requires of us that we seek to imagine how we would wish to be treated were we in the other's stead – and then that we actually go on to act towards the other in that way. (Loegstrup 2007: 85f.)

Loegstrup aside, it strikes me as an *extraordinary* omission from mainstream analytic moral theory that it should have found so little to say either about the moral significance of being a viewpoint on the world, or about imaginative identification in the deeper sense just sketched; especially given the importance that ordinary ethical thinking attaches to both these notions. It is true that at least one serious analytical ethicist, Harry Gensler, has developed a moral theory based explicitly on the Golden Rule; it is also true, of course, that Stephen Darwall continues to develop his remarkable studies of second-personality in ethics (Gensler 2013, Darwall 2006). But Gensler and Darwall and their few fellow-workers are the exceptions. By and large, this looks like a point—*another* point—at which it is perfectly reasonable for us to say "If mainstream moral theory has no place for these resources, then so much the worse for mainstream moral theory" (cp. Chappell 2014b passim, but esp. chapter 7).

Of course, the boot might be on the other foot. If (some version of) mainstream moral theory is right, and (as is usually insisted) not only right but *exclusively* right, then ideas that do not fit the true moral theory will have to be jettisoned. So, in coming to believe the true moral theory, we might lose our current intuitive grip on the moral notions of viewpoints-on-the-world, and of imaginative identification, simply for reasons of moral-theoretical parsimony.

Or we might lose that grip for another reason (instead, or as well): these notions might turn out to be *metaphysically* untenable. I've already sketched one way in which that might happen with the notion of viewpoints on the world. That the notion of imaginative identification is also a source of metaphysical puzzles seems even more obvious, though here too there has been surprisingly little exploration of the puzzles.[10]

One quick way into the puzzles that arise here is that very ordinary advisory locution "If I were you". When I tell you what I would do "if I were you", you may naturally ask how far I expect you to take the proposed identification; since if I were *literally* you, then it is just trivial that I would do whatever you do in fact do. If we are to avoid this sort of triviality the scope of my identification with you has to be weaker than that. Perhaps it will be as restricted as a game of cards; maybe all I mean to talk about when I use "if I were you" is what card I

10. One recent exception to this generalisation is Adam Morton's wonderful study *Emotion and Imagination* (Morton 2013).

would play next if I were sitting where you are now sitting, playing the bridge hand that you are now playing. In other cases, however, a wider and deeper identification seems possible. I may think about what I would do, or (wider) how I would feel, if I were Anna Karenina having just seen Vronsky's fateful riding accident; or I might wonder what it feels like to be Darcey Bussell, having just danced *Swan Lake* to triumphant applause at Covent Garden. This kind of imagining is (as Adam Morton and others point out) central to our experience of fiction. It is also central to the practice of imaginative identification that I have been wanting to say is so morally important.

The puzzle is that, at the limit, the imaginative identifications seem to be *complete*. I am not merely imagining myself, Sophie Grace Chappell, occupying the space or the clothes or the name or even the body that Anna or Darcey is actually occupying. I am imagining that *I*, Sophie, *am* that particular woman. Does that mean: my mind in her body? Not necessarily, because I can imagine, to a greater or lesser degree, my having "her mind" *as well*. (Or what we classify as "her mind" when we divide it off from her body; the difficulty of making any such division at all cleanly is one of the main reasons why, for my own part, I find it difficult even to understand what mind-body dualism is supposed to mean.)

Maybe part of what I am imagining when I imagine "being Anna Karenina", or "being Darcey Bussell", is that I share Anna's moody psychology, or Darcey's stoical ability to screen off pain in her toes while on her pointes. Maybe part of what I am imagining is even that I share Anna's or Darcey's *phenomenology*. But by the time we get to that case, the following question really does get pressing: just what am I supposing when I suppose that *I* am Darcey Bussell? What is added to the picture—which is after all, *ex hypothesi,* a picture *of Darcey Bussell*—by adding *me* to it? As Bernard Williams puts the difficulty in his remarkable and seriously undervalued paper "Imagination and the self":

> [S]uppose I conceive it possible that I might have been Napoleon—and mean by this that there might have been a world which contained a Napoleon exactly the same as the Napoleon that our world contained, except that he would have been me. What could be the difference between the actual Napoleon and the imagined one? All I have to take to him in the imagined world is a Cartesian centre of consciousness; and that, the real Napoleon had already. Leibniz, perhaps, made something like this point when he said to one who expressed the wish that he were King of China, that all he wanted was that he should cease to exist and there should be a King of China. (Williams 1976: pp. 42-3)

Now I think that in this passage Williams' argument, if that is the word for it, is plainly specious. First, his use of "Cartesian" as just quoted is lazy abuse; to think that a conception of centres of consciousness can legitimately be condemned, simply by pointing out that that conception is Descartes', is unworthy of a philosopher. (Even if the attribution is accurate—which can also be doubted.) To distance myself from this cheap *ad hominem*, from here on I shall replace Williams' "centres of consciousness" by reintroducing my earlier term, and calling them viewpoints on the world. Secondly, the fact that the real Napoleon already has a viewpoint on the world does not imply, as Williams seems to insinuate, either that there can be as it were "no room in his head" for my viewpoint on the world; nor that if my viewpoint on the world *were* to get into his head, then it would simply become the same thing as *his* viewpoint on the world.

On the intuitive conception, the answer to my earlier question "How complete can imaginative identifications be?" is then this: in cases of the most complete imaginative identification, when I imagine myself as Darcey Bussell, what I do is imagine Darcey's character and past and location and body and psychology and mood and phenomenology *with me, Sophie, not her, Darcey, as the viewpoint on the world from which all of these are experienced*. The limit of my imaginative identification is the limit of my viewpoint on the world. It is possible, on the intuitive conception, for me to go all the way *to* that limit without my simply becoming Darcey, in the way that Leibniz's interlocutor, according to him, simply became the King of China. It is only when I go *beyond* that limit, when it is not merely my body or my location or my psychology etc. that becomes hers, but my viewpoint on the world too, that I simply melt into her. So the intuitive conception has a very clear answer to my other question, "What is added to the picture by adding *me* to it". Its answer is, simply, my viewpoint on the world.

Now Williams perhaps thinks that this intuitive conception of how imaginative identification works, and of how viewpoints on the world are involved in it, is incoherent. There is certainly something deeply puzzling, perhaps even fishy, about the intuitive conception; it does seem to reify the notion of a viewpoint in a way that we might reasonably want to query further, and perhaps more searchingly than we can by simply labelling it, as I suppose we might, 'Cartesian'. I don't wish to attempt that querying here. I merely want to observe, first, that if there is something wrong with the intuitive conception, then Williams does not here succeed in locating the problem. (How could he, just by throwing the word 'Cartesian' at it?) And secondly—to return to my home key—that if this intuitive conception *is* incoher-

ent, then that is a metaphysical problem with obvious ethical implications. *If* imaginative identification is, as I have suggested, a key moral resource for us; and *if* the possibility of ethically serious imaginative identification hangs upon the possibility of viewpoints on the world working as the intuitive conception describes; and *if* that description is itself of doubtful coherence; then metaphysics poses a serious threat to a key moral resource for us. Here then is another way in which metaphysics bears very directly on ethics.

For the record, my own view is that though Williams does not make his case out properly in this passage, he is right on the larger issue: imaginative identification so conceived *is* incoherent, because it *does* involve a fallacious reification of the notion of a viewpoint on the world. That means that the intuitive conception of imaginative identification is mistaken. It does not, fortunately, mean that imaginative identification is impossible; it is just not possible *in the way we intuitively think it is*. My own suggestion about how it is possible is that imaginative identification does not work by locating a single level, the level of the viewpoint on the world, and transferring that and that only completely from one person's context to another's. Rather, it works by shifting back and forth between a lot of different levels of identification of persons, *none* of which is a level at which anything is completely transferred from one person's context to another's. So we build up an *overall* picture of how things are for someone else out of a series of *partial* views; rather as a car-driver whose windscreen is dirty, or whose vision is obscured by dazzling lights, can work round these visual obstacles by moving her head back and forth, so as to get a range of different fragmentary perspectives on what it is not possible for her to get any single complete perspective on.

However, I cannot develop this suggestion here. We have seen how the case of imaginative identification is a second case, alongside and closely connected to the case of viewpoints on the world, where metaphysical arguments can have a decisive effect on what resources are available to us for undeluded moral thinking. It is time for me to turn to my third example of a route from metaphysics to ethics via philosophy of mind.

III. PSYCHOLOGICAL CONCEPTS AS ETHICALLY THICK CONCEPTS

Suppose, if you don't mind supposing something rather far-fetched, that eliminative materialism (EM) is right in the crudest and most extreme possible version of the view. (I am not saying, nor do I need to, that any serious present-day philosopher actually accepts this version of

EM.) On this view of the metaphysics of mind, there just are no true assertions at all to be had about the desires, beliefs, and experiences either of myself or of other people. The simplest reason why that might be true would be because there *are* no desires, beliefs, experiences, or other folk-psychological phenomena; so all assertions invoking or involving them are void through false presupposition. Then for a wide range of familiar moral questions, for instance "How would you like it if someone did that to you?" or "How do you think that will make her feel?" or "Do you think that she believes that he really owes her that?", there will be no distinction between correct and incorrect answers; *any* answer we give to them will be false. Thus EM—at least in the crass and extreme version I'm considering here—will deprive us of a crucial tool for moral reasoning: the ability to project imaginatively into the emotional and psychological states of others. It will deprive us of this by the very extreme and drastic route of denying that there *are* any such emotional and psychological states.

I start with this drastic version of EM for the sake of colour and quickening of the pulse; but it should be obvious that less extreme and less implausible versions can have an importantly similar destructive or subversive effect on our intuitive methods of interpersonal moral reasoning. Less drastic and more plausible forms of EM do not say that we have *literally nothing* in our heads. They say that there is an important difference between **what we think we have in our heads**, which we describe using the dubious and incoherent folk-psychological concepts that we have inherited from our tradition; and **what we really have in our heads**. The latter is what our folk-psychological descriptions go proxy for, but those descriptions do not capture it with full accuracy; in many cases they may not be even approximately accurate. Hence we need to replace them with a more fully correct and more scientific psychology of the mind. (Perhaps we have already seen an example of just this sort of inaccuracy, in the case of the intuitive conception of a viewpoint on the world.)

The exact details of that better psychology are pending, and not my main concern here. What interests me is the following four points. First and most obviously, this is indeed a third case where our ethics is going to be decisively affected by our metaphysics. The correct account of our minds/intentional psychologies is not going to be merely *important* for what we should say about the possibility of imaginative projection; it is going to be decisive. If in fact imaginative projection *is* a possibility; and a brief scan of contemporary work in neuroethics suggests that that is by no means a foregone conclusion, and also that there is more than one way in which, in

principle, it might prove to be either impossible or a very different thing from what we currently take it to be.

Secondly, notice the speed with which EM, when thus toned down, moves from looking obviously and absurdly false to looking obviously and unexcitingly true. *Of course* we need to eliminate false and misleading components from our interpersonal psychology, and replace those dud components with notions that have the backing of our best general understanding (which obviously means, in large part, our best science). On reflection, we begin to wonder why this point should be thought proprietary to *EM*. Surely anyone is bound to say that we need interpersonal psychological concepts that get some traction on reality. Surely anyone is bound to admit that it can't be taken for granted that the concepts we currently have *do* get perfect traction on reality. This doesn't sound like eliminative materialism; it just sounds like common sense.

Thirdly, notice my use of the word "need" in the last paragraph. I submit that the needs we are talking about here are both practical and ethical needs. It is essential for us to have a set of intentional-psychological concepts that is as good as it can be, updated wherever updates are available, in order for us to make the best possible intelligible sense of other people's (and indeed our own) behaviour. This making-intelligible is crucial to our coping ethically, and indeed to our coping full-stop. This too just sounds like common sense rather than the deliverances of any special philosophical theory such as EM.

To reinforce this impression we may add—and this is my fourth point—that not only is this updating of our intentional psychology's concepts something that, we can agree with EM, we *need* to do. What's more, it is pretty obviously something that we already *do* do. The work of psychologists and psychiatrists since Freud, and the influence of modern evolutionary psychology on our self-conception, are two obvious large-scale examples of this sort of updating; but let me consider two from among many possible examples that come at a closer level of detail.

(1) In Henry Fielding's and Charles Dickens' novels, though not curiously enough in Jane Austen's or George Eliot's, we have the supposedly explanatory intentional-psychological concept of the *feminine faint*: according to this any woman (as opposed to any man) is likely to respond to severe stress or excitement by simply passing out. This concept has been updated out of our intentional psychology, first because there is in fact nothing for it to be explanatory of. (Though people do occasionally black out under stress, it is just false that there is this difference between women as such and men as such.)

And secondly because we have a complete explanation of why that concept misleadingly appeared explanatory, to do partly with women acting as they took themselves to be expected to act, and partly with the breathing-restricting clothing that women in Fielding's and Dickens' time wore. (At least if they were women of a decent class; it is noticeable that *vulgar* women in Fielding and Dickens do not usually faint.)

(2) The Protestant concept of an "uneasy conscience", brilliantly captured in, e.g., the person of Davie Balfour in Robert Louis Stevenson's *Kidnapped*,[11] has had much of the work that it used to do redistributed among a number of other and in many cases better scientifically- or philosophically-validated concepts. One of these replacement-concepts is still, to be sure, a notion of conscience. But the modern conscience occupies a much smaller range in our intentional descriptions than it did in eighteenth- or even nineteenth-century Scotland, and key space that it used to occupy has now been taken up by other psychological concepts. We now have, for instance, the rather 1960s concept of a HANG UP, the strict psychoanalytical concept of a REPRESSION, and the street psychological concept of BOTTLING THINGS UP; and we have a broader idea related to all three concepts, the roughly sub-Nietzschean idea that at least some of our most typical feelings of guilt or bad conscience track, not anything that properly-directed moral reflection is at all likely seriously to identify as *wrongness*, but ideologically-motivated aversive conditioning to which we have been non-consensually subjected in our upbringing. (Against "cross-dressing", for example.)

What we have here is the makings of an answer to Stern's question, how our metaphysics might affect our ethics, that deals in *details*. I am probably prepared to risk the general claim that our account of mind (and so, not our account of personal identity) is the most ethically significant part of our metaphysics. Whether or not that is right, what I think my two examples show is that, the more we look at the *details* of how our account of mind influences our ethical thinking, the clearer it will get that there are no really hard lines to be drawn between scientific psychological concepts, street or folk psychological concepts, and thick ethical concepts. The three kinds of concept are, of course, distinct. But they bleed into each other. And what is more, they *feed* each other.

In which of these three categories, for instance, does BOTTLING THINGS UP belong? Surely in all three. To say that someone is bottling things up is undoubtedly to bring her within the scope of a thick ethical concept. To recognise or acknowledge that in some part of your behaviour you yourself "are bottling things up" is to

11. The running contrast between Davie's plain and earnest Edinburgh Protestantism, and the wild, almost Mediterranean Catholicism of Highlanders like Alan Breck Stewart, is one of the recurring delights of reading Robert Louis Stevenson's *Kidnapped* (1886). So, confronted in Ch.27 with the card-playing of a fugitive Jacobite chieftain, Davie responds thus: "I must have got very red in the face, but I spoke steadily, and told them that I had no call to be a judge of others, but for my part, it was a matter in which I had no clearness." "'What in deil's name is this?'" is the response; "'What kind of Whiggish, canting talk is this, for the house of Cluny Macpherson?'"

acknowledge that you're getting something importantly wrong—including *morally* wrong; to say that someone else "is bottling things up" is to say that she *ought* to recognise that she has a problem and do something about it, such as stop bottling it up.

But as well as a thick ethical concept, BOTTLING THINGS UP is obviously also a concept of our folk psychology, our everyday apparatus of interpersonal description. To say that someone is bottling things up is—as usual with a thick ethical concept—not just to say something normative about her, something that is constitutively directed towards the good, but also to describe her, to say something that is constitutively directed towards the true. BOTTLING THINGS UP is meant, too, to be (or at least to be the street/ informal correlate of) some concept of our scientific psychology: it is meant to be a concept that at least goes proxy—with at least approximate accuracy—for something important in the way "proper science" might characterise our psychology. And if these threefold connections are not in place then BOTTLING THINGS UP is (little or) no *use* to us: we trust that concept as a thick ethical concept *precisely because* we think that that concept is also a functioning part of both scientific and street psychology.

Moreover, we trust this thick concept BOTTLING THINGS UP, and we do not trust some other thick concepts of older provenance that it has replaced, precisely because we think it works better than those earlier concepts. (Perhaps the thick concepts or applications of thick concepts are, or include, STIFF-UPPER-LIPPEDNESS, SELF-DENIAL, SELF-DISCIPLINE, NOT MAKING A FUSS, and (sometimes) MANLINESS.[12]) And here "working better" is a matter *both* of adequately and non-misleadingly describing what is actually there in our psychologies, both at the level of serious science and at the level of informal interpersonal description; *and also* of giving us better moral guidance, that is, better ideas about how to see and describe our situation, and about what—if anything—to do in it.

In many cases the best thing to say, when asked whether some concept is a concept of intentional psychology or a thick ethical concept, is simply to say that it's both—and that we hope it will turn out to be a concept of serious scientific psychology too. Here then is the third way in which our metaphysics, and specifically our account of mind, bears directly on our ethics.

12. See here *The Catholic Catechism* on SELF-MASTERY.

REFERENCES

Bentham, J. (1781), *Introduction to the Principles of Morals and Legislation*. London.

Chappell, T. (1995), 'Parfit, 'R-relatedness', and the Empty Question Argument', *Philosophical Quarterly* 45: 88-92.

_____ (1998a), *Understanding Human Goods*. Edinburgh: Edinburgh University Press.

_____ (1998b), 'Reductionism about persons, and what matters', *Proceedings of the Aristotelian Society* 97: 41-57.

_____ (2003), 'Persons in time: metaphysics and ethics' in H. Dyke ed. *Time and Ethics*. Dordrecht: Kluwer. pp. 189-207.

_____ (2010), 'Ethical blindspots: why Socrates was not a cosmopolitan', *Ratio* 22: 17-33.

_____ (2014a), 'There are no thin properties', in S. Kirchin ed. *Thick Properties*. Oxford: Oxford University Press.

_____ (2014b), *Knowing What to Do*. Oxford: Oxford University Press.

Chappell, S.G. (2015), 'How Encounters with Values Generate Moral Demandingness', in M. Kuehler and M. van Ackeren eds. *The Limits of Obligation*. New York: Routledge.

Darwall, S. (2006), *The Second Person Standpoint*. Cambridge, MA. Harvard University Press.

Gaita, R. (2001), *A Common Humanity*. New York: Routledge.

Gensler, H. (2013), *Ethics and the Golden Rule*. New York: Routledge.

Kant, I. (1785) *Grundlegung zur Metaphysik der Sitten*. Konigsberg.

Loegstrup, K.E. (2007), *Beyond the Ethical Demand*. South Bend, IN: University of Notre Dame Press.

Macaulay, C. (1849), *History of England*. London.

Mortion. A. (2013). *Emotion and Imagination*. Oxford: Oxford University Press.

Nietzsche, F. (1887). *Zur Genealogie der Moral*.

Parfit, D. (1984), *Reasons and Persons*. Oxford: Oxford University Press.

Rawls. J. (1971), *A Theory of Justice*. Cambridge MA: Harvard University Press.

Sprigge, T.L.S. (1983), *The Vindication of Absolute Idealism*. Edinburgh: Edinburgh University Press.

Stern, R. (1992), 'The relation between moral theory and metaphysics', *Proceedings of the Aristotelian Society* 92: 143-159.

Taylor, C. (1989), *Sources of the Self*. Cambridge: Cambridge University Press.

Timmermann, J. & Reath, A. (2010), eds.

Kant's Critique of Practical Reason: A Critical Guide. Cambridge: Cambridge University Press.

Williams, B. (1972). *Problems of the Self*. Cambridge: Cambridge University Press.

_____ (1981), 'Persons, character, and morality' in his *Moral Luck*. Cambridge: Cambridge University Press. pp. 1-19.

WHY EXTERNALISM IS NOT A PROBLEM FOR ETHICAL INTUITIONISTS

by Philip Stratton-Lake

Philip Stratton-Lake is
Professor of Philosophy at
the University of Reading.
He works on various
topics in moral philosophy,
including metaethics, moral
epistemology and the work
of Immanuel Kant. As well
as numerous articles, he
has written the monograph
*Kant, Duty, and Moral
Worth* (New York:
Routledge, 2000), and
edited a number of volumes
on ethical intuitionism,
including an important new
edition of W.D. Ross's *The
Right and the Good*.

by Philip Stratton-Lake

Ethical intuitionists are often criticised on the ground that their view makes it possible for an agent to believe that she ought to Φ whilst lacking any motive to Φ—that is, on the ground that it involves, or implies a form of externalism. I begin by distinguishing this form of externalism (what I call 'belief externalism') from two other forms of ethical externalism—moral externalism, and reasons externalism. I then consider various reasons why one might think that ethical intuitionism is defective in so far as it involves, or implies belief externalism, and argue that these objections are unpersuasive.

I.

ETHICAL INTUITIONISM IS STANDARDLY CRITICISED on the ground that it opens up too much of a gap between our moral judgements or beliefs, and motivation—that is, on the ground that it is committed to externalism. This is not true of all ethical intuitionists. Richard Price, for example, thought that there is an internal connection between our moral judgements and motivation.[1] Nonetheless, his intuitionism and realism is thought to make his internalism unsustainable. For his realism and intuitionism are supposed to make it impossible to explain how internalism is possible.[2] Intuitionists are, then criticised either on the ground that they are externalists, or if they are internalists, on the ground that they cannot explain how their internalism is possible. I do not want to deny that intuitionists like Ross, or Prichard, are externalists, or claim that Price can explain how our moral judgements can necessarily generate motives within an intuitionistic framework. Rather, the question I want to address is this: In what way is ethical intuitionism (or any moral theory for that matter) deficient if it is externalist?[3]

Before I turn to this question, it is worthwhile distinguishing three types of internalism.[4] The sort of internalism I intend to focus on may be called belief internalism.[5] Belief internalists claim that it is a conceptual truth that if A believes that she ought to Φ in circumstances C, then she is motivated to Φ in circumstances C. This is not a view about the condition under which the content of this belief can be true—the view I refer to as moral internalism below—but is a view about what it is to *believe* this content, or to *hold* this belief (whether or not the proposition believed is true). According to belief internalism, if I am not motivated to Φ, it cannot be the

case that I really believe that I ought to Φ (although it may be the case that I ought to Φ).

Some belief internalists adhere to the weaker claim that if A believes that she ought to Φ in circumstances C, then either she is motivated to Φ in C, or she is irrational.[6] Unlike the stronger version of this thesis, weak belief internalists allow that one can sincerely and comprehendingly believe that one ought to Φ without being motivated to Φ. What they are committed to, however, is that if one is not motivated, then one is practically irrational. In what follows I shall focus on the stronger form of belief internalism for two reasons. First, because it is the rejection of this that is thought to be a problem for ethical intuitionists. Secondly, although one could plausibly argue that Ross's moral realism and Humean theory of motivation commit him to the view that one can believe that we ought to Φ without being motivated to Φ, I can think of no good argument to show that these views mean that he cannot embrace, or explain weak belief internalism. In what follows, therefore, I shall use the term 'belief internalism' to refer solely to strong belief internalism.

Belief internalism should be distinguished from what may be called reasons internalism, and moral internalism. The former has been formulated most recently by Bernard Williams,[7] while the latter is usually associated with W. D. Falk.[8] According to reasons internalism it is a necessary truth that if A has a normative reason to Φ then A has either an occurrent or a dispositional motive to Φ.[9] One has a dispositional motive if it is true that one would have an occurrent motive if one knew the relevant facts and deliberated from one's existing motivational set. This is not a thesis about the relation between beliefs and motives, but one about the relation between normative (practical) reasons and motives.

Finally, moral internalists claim that it is a necessary truth that if I ought to Φ, then I have a motive to Φ. According to this view, therefore, unless I am, or can be, motivated to Φ, it cannot be the case that I ought so to act. This view is quite independent of reasons internalism, although this is obscured by the fact that Falk was both a reasons internalist and a moral internalist. But although Falk was a reasons internalist,[10] there is nothing about moral internalism which commits anyone to following him in this. Reasons internalism is a thesis about practical reasons in general, i.e. instrumental, prudential, as well as moral reasons, whereas moral internalism is restricted solely to moral requirements. One's moral internalism might stem from the fact that one thinks that judgements which involve explicitly moral concepts such as ought, right, and wrong, can only be thought of

6. Cf. Michael Smith, *The Moral Problem*, Blackwell: Oxford, 1994, p. 61.

7. 'Internal and External Reasons', in *Moral Luck: Philosophical Papers 1973-1980*, Cambridge University Press, 1981, pp. 101-113

8. '"Ought" and "Motivation"', in *Readings in Ethical Theory*, ed. W. Sellars and J. Hospers. Appleton-Century-Crofts: New York, 1952, pp. 492-510.

9. Williams seems to think this is an analytical truth *(Ibid*. p. 109), but a reasons internalist need not be an analytical reasons internalist.

10. It is for this reason that he uses the terms 'reason' and 'motive' interchangeably *(Ibid*. p. 503, and p. 504).

as expressing our attitudes, but one might be a reasons externalist because one is a cognitivist about prudential, or instrumental reasons.

Moral and reasons internalism are also independent of belief internalism. For all that moral internalism asserts is that necessarily if I am obligated to Φ then either I am motivated to Φ, or would be if I deliberated on the relevant facts. But there is nothing about this view which entails that it must be my belief that I ought to Φ which provides my motive. It might be that I ought to Φ, but that I am motivated to Φ by a thought which lacks explicitly moral content. For example, it may be that I ought morally to help John, but that I am motivated to help him not by the thought that I should, but simply by the thought that he needs help, or (to keep the Humeans happy) simply by the thought that he needs help and the desire to help him. Here the moral internalist requirement is satisfied (I am motivated to do what I should), but the belief that I ought to do this act plays no motivational role at all. If such moral beliefs need not be what motivates if the belief is to be true, then one need not be a belief internalist if one is a moral internalist.

The same point can be made in relation to reasons internalism. For all the reasons internalist requires is that if I have a reason to Φ, I must have either an occurrent or a dispositional motive. This view does not specify that this motive must be the belief that I have a reason to Φ, let alone the belief that I ought to Φ. So even if reasons internalism is true, it will not be the case that the truth of our moral beliefs will depend upon *their* ability to motivate us.

Ethical intuitionists tend to adhere to all three forms of externalism. They think that morality, at least, involves reasons and requirements which are independent of the motivational set of the agent to whom they relate. Furthermore, Ross, at least, is (arguably[11]) committed to the view that one can believe one ought to Φ without being motivated to Φ—for one may hold this belief, but lack a relevant desire. This is generally thought to be a problem for his view. I do not think it is.

II. INTERNALISM AND MORAL INTEREST

Those who think that belief externalism is a problem for ethical intuitionists typically think that the problem is that the force of moral claims will be made to depend upon a contingent concern for, or interest in, the moral status of our actions. For example, Geoffrey Warnock writes:

> Let us concede that there are, here and there in the

11. I say 'arguably' because the view that he is committed to this is supposed to stem from the combination of his Humeanism and intuitionistic realism. But all that the Humean theory of motivation involves is that a certain belief-desire pair are necessary for motivation, and that beliefs and desires are distinct existences, or psychological states. This view does not commit one to the thought that certain desires are *independent* of certain beliefs in the sense that one could have the belief without the desire.

world, some items which have the moral properties intuitionists talk about, and some which have not: why should we care? Why does the presence or absence of these properties matter? In becoming aware that some proposed course of action is, say, obligatory, I have, on this theory, added to my information, I have come to know a truth about the world. But what has this truth that I recognise to do with my behaviour? Why should I *adopt* that course of action rather than some other? The fact that the course of action is obligatory is presumably meant to be a reason for adopting it; the fact that it would, if adopted, start on a Wednesday presumably is not. But why this difference?[12]

Nowell-Smith writes in the same vein:

A new world is revealed for our inspection; it contains such and such objects, phenomena, and characteristics; it is mapped and described in elaborate detail. No doubt it is all very interesting. If I happen to have a thirst for knowledge, I shall read on to satisfy my curiosity, much as I should read about new discoveries in astronomy or geography. Learning about 'values' or 'duties' might well be as exciting as learning about spiral nebulae or waterspouts. But what if I am not interested? Why should I *do* anything about these newly-revealed objects? Some things, I have learnt, are right and others wrong; but why should I do what is right and eschew what is wrong?[13]

There are three specific worries here. If belief externalism is true, Warnock and Nowell-Smith seem to think that I could believe that I ought to Φ, but still be left with the following questions:

(1) Why should I care about the fact that I ought to Φ?
(2) Why should I do what I believe I ought to do, rather than some other action?
(3) Why is the fact that I ought to Φ a reason to Φ?

Any answer to (1) is independent of whether or not I am, or must be, motivated to act in accordance with my moral beliefs. For questions about *why* we should care about certain things are quite independent from questions of whether beliefs about these things motivate us. Even if beliefs about such things necessarily motivate us, the question of why we should care about them would remain quite untouched. The mere fact that we are motivated by itself does not vindicate our acting in one way rather than another any more than the fact that we are naturally selfish (if this is a fact) would vindicate our selfish behaviour.

It might be argued that if externalism is true then the

12. G. J. Warnock, *Contemporary Moral Philosophy*, St. Martin's Press: New York, 1967, pp. 15-16.

13. *Ethics*, Penguin: Harmondsworth, 1954, p. 41. Part of what worries Nowell-Smith about intuitionism is that he thinks it tries, illegitimately, to derive an ought from an is. I think this charge does not stick, but do not intend to argue that here. For a discussion of this point see W. Frankena, 'Obligation and Motivation in Recent Philosophy' in *Essays in Moral Philosophy*, ed. A. I. Melden. University of Washington Press: Seattle, 1958, p. 50.

question of whether our moral beliefs are well founded could be answered, and still leave (1) to be addressed, whereas if internalism is true, then once it has been shown that our moral beliefs are well founded, there will be no further question of why we should care about moral considerations. The idea here is that all the internalist has to show is that there is reason to believe that we ought to Φ, whereas the externalist also has to show that there is reason to care about whether or not we ought to Φ. Belief internalism is freed of this extra requirement, so the argument goes, because nothing more is needed to motivate us to Φ once we have come to believe that we ought to. But why should we think that the fact that we are motivated by our moral beliefs makes any difference at all to whether or not we need to offer a justification for our interest in morality? This argument does not get around my previous objection that questions about *why* we should care about something are quite independent of whether such beliefs motivate us. It might be that if belief externalism is true, then we may need some way of motivating those who are unmotivated by their moral beliefs. But this is not a normative issue, but simply a pragmatic one.

The same points can be made in response to (2). (2) asks why we should do what we believe we ought to do, rather than some other action. Suppose, for the sake of simplicity, we limit this question to cases in which our belief is true. (2) may seem like a deep question if one thinks, as Kantians tend to, that the reason why one ought to Φ is because one ought to. My view is that this cannot be a normative reason why one ought to Φ. The normative reasons why I ought to Φ are those which would constitute a correct answer to the question 'why ought I to Φ?'. But one cannot answer this question by replying, 'because you ought to Φ'. The fact that I ought to Φ cannot, therefore, be a reason why I ought to Φ. If this is correct, then (2) may be easy to answer. The reason why you should do what you believe you ought to do may simply be the reasons on the basis of which you believe you ought to do these various acts. Suppose you believe you ought to Φ, and that you believe this because you promised your friend that you would Φ. Why should you do what you believe you ought to do here? A naïve reply may be 'because you promised your friend that you would do this'. Suppose this is the correct answer when you are motivated to Φ. Does it cease to be the correct answer if for some reason you are not motivated to Φ? This may be the case if reasons internalism is true. But then it is not belief externalism, but reasons internalism which is the cause of the problem. So (2) is a question which must be answered by both belief internalists and externalists alike. Belief externalists are no worse off in this respect.

According to (3), the question which belief externalists leave unanswered is why the fact that I ought to Φ is a reason to Φ. But given what I have just said, this question does not need to be answered, either by belief externalists or internalists. For, since the fact that I ought to a) cannot be a reason why I ought to Φ, there is no question of why it is a reason.

These three specific worries may be thought to be a bad expression of a general worry about belief externalism. The general worry the passages from Warnock and Nowell-Smith express may be that belief externalism makes the normativity of moral claims dependent upon contingent interests and concerns we happen to have. If I lack a suitable concern for the morality of my actions, if the belief that I ought to Φ just leaves me cold, or unmoved, then morality will have no grip on me.[14]

But this seems simply to confuse motivating grip, with normative grip, or motivating force, with normative force. It may be the case that if I believe that I ought to Φ, but am completely unconcerned with the morality of my actions, then this belief will lack motivational force. But why should we think that because of this it will lack normative force?

Christine Korsgaard has attempted to answer this question by claiming that the normativity of moral judgements consists in two elements. They must, she maintains, be both binding and motivating.[15] Consequently, if our moral judgements are always to be normative these judgements must always motivate. If this is correct, then the scope of the authority of morality may be limited if belief externalism is true. If Korsgaard is correct, then the problem with the sort of belief externalism intuitionists like Ross adhere to is that it fails to guarantee that the moral judgements of rational agents will be normative for them. For if they lack a suitable desire, their moral judgement will not motivate them, and hence will not be normative (for them).

What, then is it for a belief, or judgement to be normative? It could be understood as follows. A normative belief is one the content of which is a normative proposition. What, then is it for a proposition to be normative? I do not wish to argue for any particular account, but something like the following may be correct. A normative proposition may be one which is either atomic or complex. A complex proposition is one which contains more than one atomic proposition, i.e. more than one truth-assessable component. An atomic proposition is one which cannot be broken down into distinct truth-assessable components. If it is atomic, it is normative if it contains a normative concept, such as ought, right, reason, etc. If it is complex, it is normative either if each

14. Christine Korsgaard makes the same point about a realist account of prudential imperatives. 'We must still explain', she writes, 'why the person finds it *necessary* to act on these normative facts, or what it is about *her* that makes them normative *for her*. We must explain how these reasons get a grip on the agent' ('The Normativity of Instrumental Reason', in *Ethics and Practical Reason*, G. Cullity and B. Gaut (eds), Oxford University Press: Oxford, 1997, p. 240).

15. 'Kant's Analysis of Obligation', p. 43.

of its components contains a normative concept, or if it entails an atomic normative proposition. But if normative beliefs are understood along these lines, then the question of whether some moral belief is normative for the individual who holds it cannot depend upon whether or not the belief motivates that agent. For, its being normative will depend upon the content of the belief, and the content of a belief is not determined by whether or not it motivates.[16] It may be that the question of whether or not someone could be said to *hold* such a belief (content) will depend upon whether or not it motivates. But if one can only be said to *hold* a normative belief if it motivates, then externalism is false. As I have already said, I am not trying to argue that belief externalism is true, but only that we would lose nothing of moral significance *if* it is true. One cannot argue against *this* with reference to an argument which entails that belief externalism is false.

The above criticism can be avoided if we think of what it is for a belief to be normative in a different way. Suppose that the normativity of a belief consists not in a normative concept being included in all of the elements of its content, but in the truth of such a content. On this account of the normativity of moral beliefs, the belief that I ought to Φ is normative, not in virtue of containing the normative concept 'ought', but in virtue of its being true that I ought to Φ If this is correct, then Korsgaard's claim that the belief would not be normative for me if it did not motivate me would amount to the claim that it would be false, if it did not motivate me.

But why should we think that our moral beliefs can only be true, if they motivate us. It couldn't be the case that I who hold such beliefs think they are false, even if I didn't care about the morality of my actions. So despite my lack of interest in morality I, at least, would think that I really ought to Φ. Why think that I must be mistaken?

Korsgaard does not answer this question. But perhaps the thought is that the fact that this belief does not motivate me falsifies the belief. For, it might be argued, if this belief did not motivate me to Φ, it *could* not be true that I ought to Φ —that is, it could not be the case that what I take to be a normative, or rational, constraint on my action, is in fact such a constraint. But this is by no means obvious. Whether or not it is true seems to depend not on belief externalism at all, but on moral internalism. For we have now moved from a thesis about the *holding* of a moral belief and motivation, to one about the *truth* of the belief and motivation. But then it is not belief externalism, but moral internalism which is the cause of the problem here. If moral externalism is

16. This argument does not depend upon the specific account of a normative proposition outlined. It depends only on the view that the normativity of a belief is determined by the normativity of its content.

true, then the truth of our moral beliefs is independent of whether or not we are motivated to act in accordance with them, and *a fortiori,* is independent of whether *they* motivate us to act in accordance with them. Consequently, there would be no reason whatsoever to think that my belief that I ought to Φ must be false simply because it does not motivate me (not even dispositionally). There is, therefore, no reason to think that if belief externalism is true, that it is in virtue of this that, if I am unconcerned with the morality of my actions, my moral beliefs cannot have normative force. If they lack normative force in the sense that their content would be false this will not be because of belief externalism, but because moral internalism is true.

The thought behind Korsgaard's claim that moral judgements can only be normative if they motivate may be that moral judgements are supposed to guide our action, and they cannot guide action unless they can motivate us to act in the appropriate way. But this argument plays on an ambiguity about something's being action guiding. The notion of being action guiding can be understood either empirically or normatively. To say that a belief is action guiding in the empirical sense is to say that it does, as a matter of fact, guide someone's action. To say that it is action guiding in the normative sense is to say that it *ought* to guide their action. Absent some implicit commitment to moral internalism, the view that moral beliefs must motivate an individual for them to be normative for her stems from the failure to distinguish the empirical and normative sense of being action guiding. For clearly something cannot guide us in the empirical sense unless it motivates us. It could, however, be true that our moral beliefs ought to guide our deliberation and action, even when they do not motivate us.

The view that our moral beliefs must motivate us in order for them to be normative for us may be thought to follow from the 'ought' implies 'can' principle. The idea may be that we can only be required to Φ if we are able to Φ, but we will only be able to Φ if we can be motivated so to act. It may then be claimed that the problem for belief externalism is that it does not guarantee that everyone can be motivated to act morally and hence can be morally required to act in this way.

I do not think this is a very compelling argument. Even if one accepted that the 'ought' implies 'can' principle implies that if one cannot be motivated to Φ then one cannot be morally required to Φ,[17] belief internalism would not guarantee that everyone would be able to be motivated to Φ. It could only do this if it were possible to persuade everyone who ought to Φ *that* they ought to. But it may be the case that there are some—perhaps

17. There is reason to think that the 'ought' implies 'can' principle does not imply anything of the sort. One may think, as Sinnott Armstrong does, that the form of implication at work here is not strict entailment, but merely conversational implication ('"Ought" Conversationally Implies "Can"', *Philosophical Review,* vol. XCIII, 1984, pp. 249-261). Alternatively, one might think, as Barbara Herman does, that what the principle requires is not that we be able to Φ here and now, but that Φ-ing is a type of act that we can perform ('Obligation and Performance', in *The Practice of Moral Judgment,* Harvard University Press: Cambridge, Massachusetts, 1993, p. 163-4).

many—who could not be convinced that they ought to Φ; and if that is the case, then there might be some who could not be motivated to o) even if moral belief internalism were correct. Someone might have certain religious convictions which meant that they could not even contemplate Φ-ing, let alone come to believe that that is what they ought to do. More generally, someone might be completely convinced that amoralism is true, and thus could not be brought to believe that they are morally required to do anything. In such cases even if it were true that if one came to believe that one ought to Φ one would be motivated to Φ, it might not be the case that one can be motivated to Φ, because one might not be able to be brought to believe that this is what one ought to do. So if the ought implies can principle means that belief externalism represents a threat to the authority of morality in the suggested way, moral belief internalism would not alleviate this threat. So far as I can see, therefore, there is no reason to think that *de se* deontic beliefs, beliefs about what I should do, can only be normative if they motivate.

III. BELIEF INTERNALISM AND MORAL PLATITUDES

According to one line of argument the problem with belief externalism is that it cannot accommodate certain platitudes. For example, Falk argued that belief externalism cannot capture the following platitudes:

(a) That 'we commonly expect that in thinking ourselves obliged we *ipso facto* feel some constraint to do what we think we ought to do'.
(b) That 'when we try to convince another that he ought to pay his bills, we expect our argument if accepted to effect some change of heart in him'.
(c) That 'we should think it odd to receive the answer: "Yes, I know now *that* and *why* I ought to pay my bills, but am still without any incentive for doing so."'[18]

But it is by no means obvious that belief externalists like Ross could not accommodate these platitudes. For these platitudes do not express views about how we *must* be motivated if we are to hold moral beliefs, but how we typically are, or expect other to be, motivated.[19] The belief externalist can explain this simply by pointing out that most of those who hold moral beliefs, or who take the trouble to deliberate on the morality of their actions, are, as a matter of fact, typically concerned about the morality of their actions. This is all that is needed to accommodate the general expectation that our motivations can generally be expected to track our moral judgements.

18. 'Obligation and Rightness', *Philosophy*, XIX, 1944, pp. 139-41.

19. See Frankena, *Ibid*, p. 68.

Michael Smith, however, has recently argued that the belief externalist's attempt to do this gives rise to a totally unacceptable picture of the morally good and strong willed individual.[20] For, he argues, the belief externalist can only explain how the motivations of such individuals track their moral judgements by attributing to them a de dicto desire to do the right thing. But this gives rise to a bizarre conception of the good person, as someone who only cares nonderivatively about one thing—doing the right thing, whatever that is. A morally good person will not care for the well-being of others, justice, equality, fidelity, etc., for their own sake, but only as a means of satisfying his desire to do the right thing. The externalist can thus explain the reliable connection at issue only at the cost of elevating a moral vice to the status of the one and only moral virtue. If it is true that belief externalists can only capture the platitude that good and strong willed people will be motivated to do what they think is right at the cost of turning moral virtue into a kind of moral fetishism, then any moral theory which is committed to belief externalism will certainly be the worse for it.

Let us concede that externalists like Ross can only explain the reliable connection between the moral beliefs and motivations of morally good and strong willed individuals by attributing to them a de dicto desire to do whatever is right. Does this entail that all of their other concerns and desires are derived from this desire? The only reason I can see for thinking that is true is if we assume that the belief externalist is committed to the view that the *only* nonderivative desire a morally good person will have will be the de dicto desire to do the right thing, and that her only reason for doing anything will be as a means of satisfying this desire. But there is nothing in the belief externalist's view which commits him to that! Externalists like Ross can quite sensibly allow that the de dicto desire to do what is right is simply one of the nonderived desires a good person will have. She will no doubt also have nonderived desires to promote the welfare of her family, friends, and others as well, and one might reasonably assume that a de re desire to do what she believes is right might be amongst these. The de dicto desire to do the right thing is assumed merely to explain why, when she abandons her belief that Φ-ing is right, and comes to believe that Ψ-ing is right, she will be motivated to Ψ. But this is compatible with the view that she might have some other nonderivative desire which meant that she is still motivated to Φ even though she no longer thinks that it is right to Φ. Indeed, in cases of conflict, say between the demands of morality and one's attachment to a loved one, one would expect morally good people to be motivated in his way. Suppose I am a morally good person, and know that my son has

20. *The Moral Problem*, chapter three.

murdered someone. Suppose further that I know that if I turn him in, he will receive the death sentence. I might begin by thinking that I ought not to turn him in, that despite the terrible thing he has done, this is too much to ask of me. But suppose I eventually come to the view that I really should turn him in, and act accordingly. As I have said, it is quite compatible with belief externalism, and with the view that morally good people will have a de dicto desire to do the right thing, that they will also have other nonderived concerns, such as a concern for their loved ones. In the case I have described, the de dicto desire to do the right thing will explain why, when I come to believe that I ought to turn my son in to the police, I will be motivated to do so. But my underived concern for my son will mean that I nonetheless desire, and am motivated, not to turn him in.[21] The reason I do turn him in is because my de dicto desire to do what is right wins out (because I am strong willed). It is true that in this case I lack a de re desire to do the right thing. I only have a derived desire to turn my son in to the police. But this is how it should be. As Hallvard Lillehammer points out, it would be a form of moral fetishism to assume that in situations like this a morally good person would, as Smith claims, have a de re, but not a de dicto, desire to do the right thing.[22]

Once we note that the belief externalist's view that good people will have a nonderivative de dicto desire to do the right thing does not rule out the possibility that they will have other nonderivative desires, and is consistent with the view that they will often, but not always, also have a de re desire to do the right thing, there is no reason to suppose that the belief externalist can only accommodate moral platitudes at the cost of a deeply unattractive account of the morally good person.

There are, no doubt, other objections to belief externalism which I have not considered. But I believe I have dealt with the most common and important objections, and conclude that critics have failed to show that ethical intuitionism, or any other moral theory for that matter, is the worse for its commitment to belief externalism. This is not to say that the debate about belief externalism is unimportant. One might think that the debate about belief externalism is important because if one could show that belief externalism is true, it would make any form of noncognitivism in ethics deeply implausible. My aim has simply been to show that a moral theory is none the worse for being committed to belief externalism.

21. I think cases like this cast doubt on John McDowell's thesis that in perfectly virtuous individuals, considerations which conflict with their moral verdicts are silenced ('Virtue and Reason', *The Monist*, 62, 1972, p. 335).

22. 'Smith on Moral Fetishism', *Analysis*, 57, 1997, p. 192.

WHY EXTERNALISM IS A PROBLEM FOR ETHICAL INTUITIONISTS

Commentary by Jimmy Lenman

JAMES LENMAN grew up in Dundee and attended university at Oxford and St Andrews. He has been a lecturer at Lancaster and Glasgow and a Fellow at the Center for Ethics and the Profession at Harvard. He is now a professor of philosophy at the University of Sheffield. He has published papers about, inter alia, expressivism and quasi-realism, consequentialism, moral responsibility, moral epistemology, moral error theory and fictionalism, moral naturalism, moral objectivity, practical reason, motivation and agency, the ethics of imposing risks and the significance of death, human extinction and work. He is very keen on walking in the British countryside, reading old books, food and watching snooker on the telly.

WHY EXTERNALISM IS A PROBLEM FOR ETHICAL INTUITIONISTS

Commentary by James Lenman

1. *Confessions* 11, 14, 7. ("For what is time? If nobody asks me, I know. Challenged to explain, I am baffled.")

IF THERE ARE ANY THINGS PHILOSOPHERS ARE FOR, one of those things is surely helping us to see when things that seem obvious and straightforward are really not obvious and straightforward at all. Numbers, for example, do not seem terribly perplexing or mysterious when we are applying the elementary arithmetic we learned at school to checking our change in the supermarket. But when we get started thinking in earnest, numbers are pretty queer customers. And so it is with many topics: the self, modality, time. Augustine's expression of perplexity remains perennially apposite: *Quid est enim tempus? Si nemo ex me quaerat, scio; si quaerenti explicare velim, nescio.*[1] It is, in general, I take it, a strike in favour of philosophical theories when they help to tame such mysteries; a strike against them when they make them darker.

What is it to be normative?, asks Philip Stratton-Lake (1999: 317-8). And he offers a sort of answer. A normative proposition is one which contains a normative concept, such as *ought*, *right* or *reason*. Of course that only helps us much if we understand *those* concepts. It is a vexed and contested question how we should do so. Some naturalists offer reductive analyses that seek to capture the meaning of normative terms in ways that employ only non-normative concepts. Expressivists reject such reductive aspirations but seek to secure a respectable place for normative concepts in a broadly naturalistic take on the world by characterizing the fundamentally non-cognitive states of mind we employ them to express. Ethical intuitionists also reject any reductive aspirations, taking normative (and, in particular, moral) concepts to denote non-natural properties resisting definition or analysis which somehow make themselves evident to us.

For expressivists and at least some reductive naturalists there may be a close connection between moral and other normative concepts and motivation. Being motivated to ϕ need not, I take it, imply that one does ϕ but does, I take it, imply that one at least to some extent is disposed to or desires to ϕ. For expressivists, believing one ought to ϕ in C just is, perhaps *inter alia*, a state of being motivated to ϕ in C. And for at least some naturalists, believing that one ought to ϕ in C is to believe something like e.g. that one's ϕ-ing in C would in

some way speak to or further some desires or goals that one has. [2] The former, expressivist picture, at least, and maybe some versions of the latter naturalist one are evidently hospitable to the idea of a close such connection that holds from conceptual necessity, of the kind that Stratton-Lake takes to be definitive of belief internalism. (1999: 312-13)

Ethical intuitionists who, like Stratton-Lake, are belief externalists sever any such connection between normativity and motivation. To their opponents this can seem baffling. Why, both G. J. Warnock and P. H. Nowell-Smith wonder, in passages quoted by Stratton-Lake, if moral facts and properties are what ethical intuitionists think they are, should anyone care about them?[3] But Stratton-Lake thinks this question has nothing to do with whether moral facts and properties motivate us. It is just another normative question, inside the normative game. Motivating force and normative force are entirely distinct. Warnock and Nowell-Smith are both quite at a loss to see what, if moral facts and properties are what rational intuitionists suppose, their practical significance can possibly be. But Stratton-Lake takes the essential practical significance of moral facts and properties to reside in their normative significance. He rejects (1999: 317-8) Christine Korsgaard's contention (1996: 52) that normativity must make some connection with motivation by insisting, as I noted just above, that for a proposition to be normative is just for it to contain some normative concept, *reason*, *ought*, or whatever. So ethical intuitionism has no problem with externalism.

To see why this is wrong, consider someone I call Matilda.[4] Matilda, let us suppose, is a police officer. And she is, let us suppose, just the sort of person we would all want to find in the police force. If we knew her and if we believed her a typical member of the police force, we would feel very comfortable and very safe, not just from crime but from police brutality, corruption or wrongful arrest. For Matilda hates crime, hates corruption and could never live with herself if she ever thought she had been instrumental in imprisoning an innocent person. She hates and is disgusted by all forms of dishonesty, all forms of violence, every kind of abuse of power. She wants with great passion to do what she can, within the law, in her professional life, to combat these things and to be a source of help and succour to those who stand to suffer as a result of these things.

And Matilda is more than just a model police officer. Her friendly and sunny disposition means she is pleasant and kind to all who come in contact with her. She is a wonderful colleague. She is married to Matt to whom she is a loving and faithful wife. And she has many friends to

whom she is devoted. Matilda is what most of us would call a very good person.

Her goodness, what is more, is pretty robust under the various kinds of pressure to which goodness is subject. It is important to her to treat those around her well, always to be warm and loving to her family and friends, always at least respectful and civil to those she deals with professionally, be they her colleagues or those people, not always easy to like, whom she has to arrest and hold in custody. Most of the time this is easy. She has a sunny, sociable nature and making herself agreeable to others is second nature for her. But even when her sunny, sociable nature is in abeyance, as it occasionally is, for the best of us have our bad days, she manages to behave in these ways simply because that is important to her. In some measure, perhaps, because she is in the grip of certain ideals of character that so prescribe.

What might that mean? Well, we might understand ideals of character in terms of what we will call systems of norms concerning themselves with how a person should live her life. By a system of norms I mean just a set of commands or prescriptions as these might be expressed by sentences in the imperative mood. One such system might be NB. NB might express an ideal of character that permits suspension of civil and agreeable engagement with those around one at times when one's mood so disposes one. NG, on the other hand does prohibit these behaviours as well as e.g. thinking them appropriate targets for negative reactive attitudes and other sanctions, prescribing the raising of one's children to be disposed to refrain from these things, and so on. And of course I want to say the explanation of Matilda's very robust disposition to fidelity and friendliness is a consequence of her acceptance of an ideal of character which is of a kind with NG and not with NB. To accept an ideal of character is not here to be understood in terms of *believing* something to be true about it. Rather it is a matter of liking it, or preferring it to others, of liking and admiring people who live in the way it prescribes and of desiring to be a person of that sort. It is a matter, in short, of how one is motivated.

But why might it be that Matilda embraces the ideals of character she does? Why does she embrace the relatively agreeable ideal expressed by NG rather than the relatively disagreeable one expressed by NB? Well, part of the explanation may be that she also embraces certain what we may call *ideals of community*. These are to be understood in terms of systems of norms for the governance of human moral communities. These are again just sets of imperative sentences recommending that human communities organize themselves in certain ways and

not in others, norms perhaps that recommend solidarity and mutual helpfulness and concern as opposed to mutually indifferent individualism and deep-seated sectarian division. There are some such ideals that Matilda embraces and these constrain the kinds of ideals of character she embraces just as the ideals of character she embraces constrain the particular courses of action she favours or disfavours.

So the ideals of community Matilda embraces constrain the ideals of character she embraces. And the ideals of character she embraces constrain the particular actions she chooses to perform. We could perhaps here again introduce some more terms of art. We could describe her deference to ideals of community in her choice among ideals of character by saying she is *principled* and we could describe her deference to her ideals of character in choosing particular actions by saying she has *integrity*. Among her many admirable characteristics Matilda has both of these. Indeed in having both of these she is conforming to her favoured ideal of character which itself enjoins her to be this way.

I have now told you a lot, I think you will agree about what we might call Matilda's moral psychology. And I think that she looks like a rather admirable soul. But you may have noticed that in telling you what I have I have exercised what I hope is a significant self-denying ordinance. I have told you nothing whatever about her *moral beliefs*. What I have told you about is about her *desires*, broadly conceived. What she wants, what she likes, what she cares about. I have described in fair detail the passions of her soul but I have not said a word about what Matilda believes to be right and wrong.

There is good news here and there is bad news. I will give you the bad news first. Let us suppose Matilda is what I will call an *algist*. We all know what a hedonist is, from the Greek, *hedone*, meaning pleasure. Algism is an analogous position from *algos*, the Greek for pain (hence our *analgesic*). According to algism, what makes right acts right is their tendency to be productive of conditions maximally involving of pointless suffering, while acts which fail to do this, or worse still, acts which serve to diminish such suffering, those acts are wrong.

That, you will agree, looks like bad news. But it is not very bad news. This is because of what I will tell you next which is the good news. *Matilda is an amoralist*. She doesn't give a toss about what is morally right and wrong. It has, with one small qualification I'll mention in a second, zero impact on her practical thought, on her desires, aversions, loves, hates, projects and ideals at least in terms of the rigorously noncognitive characterization of ideals I set out above. So Matilda is not pre-

vented by her horrible moral beliefs from continuing to be the wonderful human being I have described thus far.

I mentioned that there was one small way in which Matilda's moral beliefs impact on her motivation and that is this. She keeps them to herself. She has noticed that other people do care about right and wrong in ways she does not. So she fears that if she shares her views on these topics she risks consequences that she will not like. For one thing she might influence people, if they are impressionable and young, to go out and act in algist ways. Because algism is so crazy that is unlikely and the greater risk is that she will cause people to think she is a monster for professing such beliefs. So she avoids sharing them. This can be hard on her as it is hard to do this without lying and she is, as we have seen, a very good and principled soul but she does her best. She is rather good at a certain evasiveness. Visiting a school, a child asks her, *Don't you think the things the Moors Murderers did were very wrong?* Rather than reply, truthfully, *No, I don't*, she will say something true along the lines of, *I hate what those people did and could not bear to live in a community where such behaviour did not attract the strongest criminal sanctions.*

That is the only practical significance Matilda accords moral beliefs. She is an amoralist. She is emphatically *not* weak-willed. She feels *no* motivation to act as her moral beliefs tell her she ought to act. The *akrates* is motivated all right to do the things he fails to do - his motivation is just not up to the job. That is why he sees his failure as a failure. But Matilda does not see her failure to be motivated to act on her moral beliefs as a failure. She embraces her highly algism-opposed motivational make-up just as Frankfurt's willing addict embraces his habit (Frankfurt 1988). She doesn't just want to act as she does, but wants wholeheartedly to do so in a way that is highly stable under reflection.

The supposed possibility of an amoralist is a favourite datum offered in support of externalism (e.g. in Brink 1989 and Svavarsdottir 1999). In the most familiar imaginings of this possibility, the amoralist is a horrible person with good but motivationally ineffectual moral beliefs. If this does represent a genuine possibility, so surely is a case like Matilda of a quite lovely person with awful moral beliefs. But reflection on a case like Matilda surely leads us to question whether we do here confront a genuine possibility. It is natural to wonder if the lesser plausibility of this more benign species of amoralist might not be due to the obvious ineligibility, in Matilda's case, of an "inverted commas" reading of her moral beliefs (c.f. Hare 1952: 1, 124-6 and 163-170).

Expressivism seems to square with the plausible

supposition that my characterization of Matilda's moral psychology makes no clear sense. When I talk about someone's moral beliefs I am really, according to expressivists, just talking about passions in their souls of just the kind I described at great length in the case of Matilda. And when, nearing the end of this lengthy description I announced that I had, up to that point, told you nothing about her moral beliefs, any expressivists among you will have shaken your heads and thought, *Oh, but you have*. This expressivist diagnosis is surely an appealing one. It might look at first glance as if you and I – assuming you, like me, are not an algist – disagree with Matilda, given that our moral beliefs seem to contradict what I have called her "moral beliefs". But on reflection this does not seem so obvious. When it comes to all practical questions, all questions of *what to do*, it seems likely that most of us have very little disagreement with Matilda at all. She does say various things about what she calls "right" and "wrong", etc. but I'm not at all confident in saying I disagree with these judgements of hers. Indeed I have no idea what she might mean by them. Disconnected as they are from Matilda's thinking about how to live her life, these concepts now seem quite meaningless.

Some naturalists might also be well placed to make sense of the difficulty I think we find in making sense of Matilda. If believing that one ought to ϕ in C is to believe that one's ϕ-ing in C would in some way speak to or further desires or goals that one has, Matilda must believe that maximally promoting pain would speak to or further some desire or goal that she has. But she does not believe that. There is no desire or goal that she has that would be furthered by doing this and there is no desire or goal that she has that she believes would be furthered by doing this. And other versions of reductive naturalism might claim an equal entitlement to baulk at Matilda. Matilda's notion that it is good that there be as much pain as possible might, in their view, make roughly as much sense as someone insisting that the sieve in which the Jumblies went to sea is an exemplary case of *seaworthiness* when the person in question is an expert sailor with the interests and motives we would expect from a good and reliable sailor and understands what "seaworthiness" means, i.e. no sense.[5]

But things look decidedly awkward for the ethical intuitionist who embraces belief externalism. For this ethical intuitionist there seems to be no reason why we cannot coherently imagine someone with the psychology I have imputed to Matilda. Indeed there is no very obvious reason why someone could not coherently *be* someone with the psychology I have imputed to Matilda. Of course Matilda makes no attempt to bring her actions

5. Peter Railton develops this analogy in his brilliant 1989.

and desires into conformity with her moral beliefs but why would coherence, *as such*, demand that any more than it demands that she bring her actions into conformity with the moral beliefs she attributes to her favourite fictional character? Of course we may think an ideal moral character is someone who does bring her actions and desires into conformity with her moral beliefs. And perhaps Matilda herself believes she ought, morally, to do this. But that is just another of those moral beliefs of Matilda's which she fails, as we have stipulated, to give a toss about.

Matilda's moral beliefs have none of the practical significance moral beliefs ordinarily enjoy. She doesn't care about them, doesn't attempt in any way to conform her actions to them. Of course, the ethical intuitionist may nonetheless insist, this does not stop them being normative. They are normative just in the way Stratton-Lake spells out for us in being beliefs whose propositional contents contain at least one normative word, a word like 'ought', 'right' or 'reason'. But that is the problem. For we can get no proper grip on what, as Matilda uses them, these words could possibly mean.

The expressivist has a story about what these words mean according to which Matilda's psychology makes no sense. She can't mean *that*. At least some naturalists have stories about what these words mean according to which Matilda's psychology makes no sense. She can't mean *that* either. But for the ethical intuitionist who embraces belief externalism Matilda's moral psychology should make sense. And yet it surely does not. For when the disconnect between these normative concepts and what motivates us is made this sharp, we lose all grip on any supposition that we understand these concepts at all. In the mouth of Matilda, e.g. the word "right" in "It is always right to cause as much pain as possible" is reduced to mere noise.

Matilda's moral psychology makes no sense. By the lights of externalist ethical intuitionism, it should make sense. So we should not accept externalist ethical intuitionism. This is a version of what I call the *disconnect argument* (Lenman 2008: 359). I think it remains in the spirit of the Nowell-Smith/Warnock point to which Stratton-Lake is responding: in effect, that the ethical intuitionist renders "ought"s as "is"s that are "normative" only in, at best, a rather dark Pickwickian sense, one that leaves them no less thoroughly disconnected from intrinsic practical significance than Hume alleged "is"s of other sorts to be.
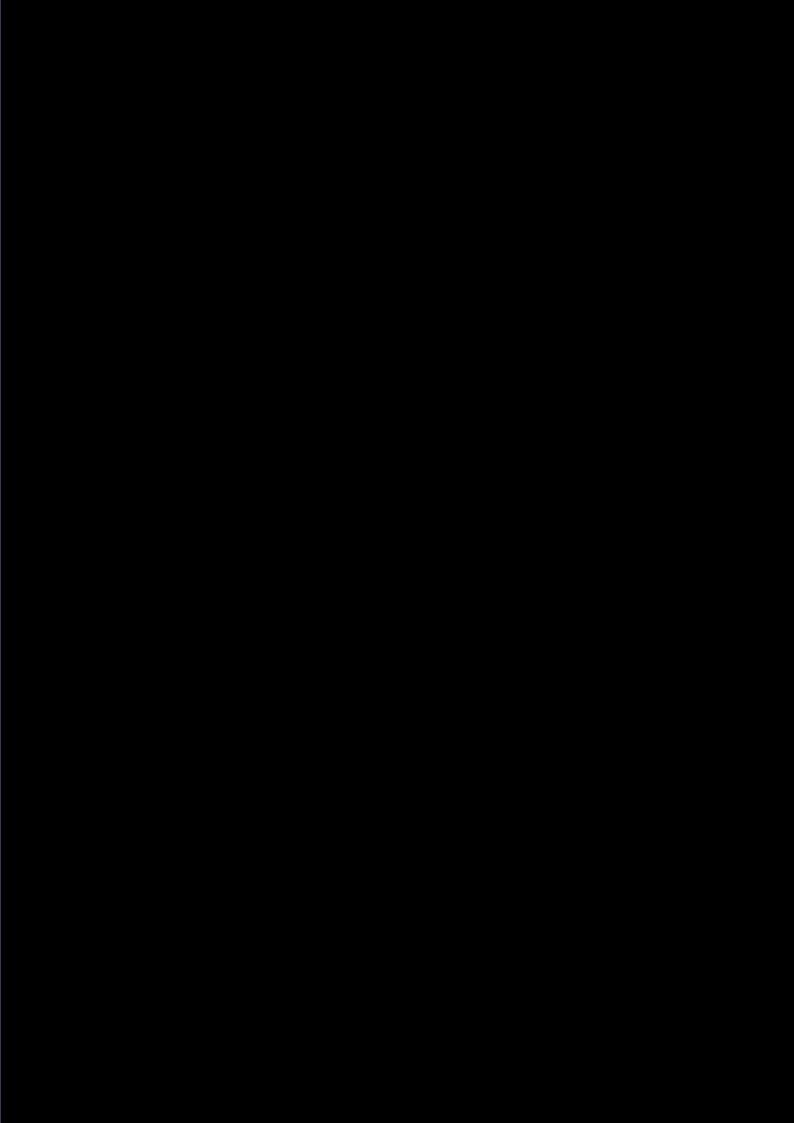
The question whether I, or Matilda, should care about the facts about what I, or she, ought to do, is, the externalist intuitionist rightly insists, a normative ques-

tion. But he cannot tell us, and if we believe his view, we cannot know, what that means. If you want to be a belief externalist, this short paper might not worry you too much. If you want to be an ethical intuitionist, it may leave you with plenty of options but if you want to be both an ethical intuitionist and a belief externalist, I think Matilda is trouble. If you want to be both an ethical intuitionist and a belief externalist, you disconnect normativity from motivation, from love and hate, aversion and desire in ways that preserve its purity only at the expense of its intelligibility. Externalism is a problem for ethical intuitionism.

REFERENCES

Brink, D. (1989), *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Dreier, J. (1990), 'Internalism and Speaker Relativism', *Ethics* 1010: 6-26.

Frankfurt, H. (1988), 'Freedom of the Will and the Concept of a Person' in his *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press. pp. 11-25.

Gibbard, A. (2003), *Thinking How to Live*. Cambridge MA: Harvard University Press.

Hare, R.M. (1952), *The Language of Morals*. Oxford: Clarendon Press.

Korsgaard, C. (1996), 'Kant's Analysis of Obligation: The Argument of *Groundwork* I' in her *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press. pp. 43-76.

Lenman, J. (1999), 'The Externalist and the Amoralist', *Philosophica* 27: 441-457.

_____ (2003), 'Moral Deviants and Amoral Saints: A Dilemma for Moral Externalism', *Southern Journal of Philosophy* 61: 223-240.

_____ (2008), 'Actions, Motives and Causes', *Philosophical Quarterly* 58: 353-362.

Nowell-Smith, H. (1954), *Ethics*. Harmondsworth: Penguin.

Railton, P. (1989), 'Naturalism and Prescriptivity', *Social Philosophy and Policy* 7: 151-174.

Stratton-Lake, P. (1999), 'Why Externalism is not a Problem for Ethical Intuitionists', *Proceedings of the Aristotelian Society* 99: 77-90.

Svavarsdottir, S. (1999), 'Moral Cognitivism and Motivation', *The Philosophical Review* 108: 161-219.

Warnock, G.J. (1967), *Contemporary Moral Philosophy*. London: MacMillan.

VI

THE VIRTUAL ISSUE

An Aristotelian Society Publication