

The Morality of Common Sense: Problems from Sidgwick



NAKUL KRISHNA
Balliol College
University of Oxford

A thesis submitted in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy

Michaelmas Term 2014

68,158 words

Acknowledgements

I am grateful above all to my supervisor, Professor Roger Crisp, for guiding this dissertation with open-mindedness and a light touch. I owe sincere thanks to my secondary supervisor, Professor Terence Irwin, and to my examiners, Professor Stephen Mulhall and Professor Martha Nussbaum for the time and sympathetic attention they gave to my thesis.

Much in the following pages is the product of conversations with my teachers in the past few years, among them Dr Chris Brooke, Mrs Lesley Brown, Dr Gideon Elford, Professor Cécile Fabre, Professor David Miller, Professor Benjamin Morison, Professor Jessica Moss, Dr Jonas Olson, Dr Matthew Shapiro, and Professor Marc Stears. The views in this thesis have been honed in seminar room interactions with Dr Daniel Butt, Professor Simon Caney, Professor John Gardner and Professor Roger Scruton. I am grateful for comments on portions of this thesis from Professor Krister Bykvist, Dr Edward Harcourt, and Professor Adrian Moore.

I came by my most original thoughts on the subject of this thesis in the course of conversations with Yuuki Ohta, Alec Hinshelwood and Paul Sagar, all of whom read and commented on my drafts. My ideas were refined in conversation with Tom Ainsworth, Luke Brunning, Elena Cagnoli, Anna Comboni, Rachel Fraser, Dan Hitchens, Christopher Jay, Will Jefferson, Claire Kirwin, David Mathers, Andreas Mogensen, Jessie Munton, Colm Ó Siochrú, Nathan Pinkoski, Mike Price, Stefan Sienkiewicz, Robert Simpson, Sarah Lane Smith, Ben Sorgiovanni, Amia Srinivasan, Damien Storey, Mara van der Lugt, and Robert Watt. I am also grateful for my conversations on general themes with Joseph Carlsmith, Gabriel Roberts, Emily Jones, Tarunabh Khaitan, Alex Bubb, Rahul Rao, and Anisha Sharma.

The material comprising this thesis was tested before a number of audiences: the DPhil Seminar, Ockham Society, Publication Workshop and Politics Work-in-Progress Seminar (Oxford), the Berkeley Graduate Conference on the History of British Political Thought (Berkeley), the Cambridge Graduate Conference on Political Thought and Intellectual History (Cambridge), and the conference on the Moral and Political Legacy of Bernard Williams (Oxford). I am grateful to the organisers of these conferences/seminars, and to my respondents and other questioners on these occasions.

I am indebted to the financial support of the Rhodes Trust and the Clarendon Scholarship for funding me for six years of study and research in Oxford. My study of Ancient Greek was aided by a grant from Exeter College. I was also the recipient of generous financial aid from Balliol College in the closing stages of my DPhil. The administrative staff both at the Faculty of Philosophy and at all three of my colleges – Exeter, Corpus Christi and Balliol – were infinitely supportive during my time in Oxford. In particular, I am grateful to the library staff who did so much to make Oxford a congenial place for solitary study.

For their support and bemused tolerance of my academic work, I thank my parents G Vishwanath and K R Saraswathi, my sister Nandini Vishwanath and my brother-in-law Vinodh Jayaram. For their part in raising me, and in different ways introducing me to the life of the mind, I thank my grandparents C S Gopalkrishna, P S Bhagirathy, K Raman and Radha Raman.

Finally, I should like to thank Anmol and Sarita Vellani, in conversation with whom I first conceived of the idea of applying to Oxford and of studying philosophy; Fram Dinshaw and Claudia FitzHerbert, in whose house the final draft of this thesis was written; and Wolsey, for his company while I was writing it.

Abstract

Much modern moral philosophy has conceived of its interpretative and critical aims in relation to an entity it sometimes terms 'common-sense morality'. The term was influentially used in something like its canonical sense by Henry Sidgwick in his classic work *The Methods of Ethics* (1874). Sidgwick conceived of common-sense morality as a more-or-less determinate body of current moral opinion, and traced his ('doxastic') conception through Kant back to Aristotle's *Nicomachean Ethics* and the practice of Plato's Socrates before him. The Introduction to this thesis traces the influence of Sidgwick's conception both on subsequent (mis)understandings of Socratic practice as well as on the practice of moral philosophy in the twentieth century.

The first essay offers a challenge to Sidgwick's understanding of Socratic practice. I argue that Socrates' questioning of his interlocutors, far from revealing some determinate body of pre-existing beliefs, is in fact a demonstration of the dynamic and partially indeterminate quality of common-sense morality. The value for the interlocutor of engaging in such conversation with Socrates consisted primarily in its forcing him to adopt what I term a deliberative stance with respect to his own practice and dispositions, asking himself not 'what is it that I believe?' but rather, 'what am I to believe?' This understanding of Socratic practice gives us a way of reconciling the often puzzling combination of conservative and radical elements in Plato's dialogues.

The second essay is a discussion of the reception of Sidgwick's conception of ethics in twentieth-century Oxford, a hegemonic centre of Anglophone philosophy. This recent tradition consists both of figures who accepted Sidgwick's picture of moral philosophy's aims and those who rejected it. Of the critics, I am centrally concerned with Bernard Williams, whose life's work, I argue, can be fruitfully understood as the elaboration of a heterodox understanding of Socratic practice, opposed to Sidgwick's. Ethics, on this conception, is a project directed at the emancipation of our moral experience from the many distortions to which it is vulnerable. Williams's writings in moral philosophy, disparate and not entirely systematic, are unified by these emancipatory aims, aims they share with strains of psychoanalysis except in that they do not scorn philosophical argument as a tool of emancipation: in this respect among others, I claim, they are fundamentally Socratic.

Contents

ACKNOWLEDGEMENTS	2
ABSTRACT	3
CONTENTS	4
INTRODUCTION	5
MUST WE SAY WHAT WE THINK? SOCRATES, SELF-KNOWLEDGE, AND COMMON-SENSE MORALITY	43
1. INTRODUCTION	43
2. 'SAY WHAT YOU BELIEVE'	54
3. WHY SAY WHAT ONE BELIEVES?	60
4. FORMING BELIEFS	67
5. THEORETICAL OR DELIBERATIVE?	80
6. SINCERITY, WILLINGNESS AND SOCRATIC PROTREPTIC	94
7. SCOREKEEPING IN A SOCRATIC LANGUAGE GAME	120
8. CONCLUDING REMARKS	146
'AN ACUTE AND WARY SENSE OF THE ORDINARY': COMMON SENSE AND CRITIQUE IN THE PHILOSOPHY OF BERNARD WILLIAMS	149
1. DESCRIPTIVE AND REVISIONARY ETHICS	150
2. WILLIAMS AND HIS PEERS	157
3. LIBERATION PHILOSOPHY?	181
4. IMPRESSIONISTIC ETHICS	208
5. STYLE AND STYLISHNESS	229
BIBLIOGRAPHY	234

Introduction

It is sometimes said, either irritably or with a certain satisfaction, that philosophy makes no progress. It is certainly true, and I think this is an abiding and not regrettable characteristic of the discipline, that philosophy has in a sense to keep trying to return to the beginning: a thing which is not all that easy to do.

Iris Murdoch¹

‘Where should we start in philosophy?’ the acolyte asks the Zen Master. ‘From where we are’, the master replies. The exchange suggests an attractive conceit: philosophy as a journey, the philosopher as traveller. Like other poetic conceits, it can be elaborated endlessly, to illuminating, puzzling or comic effect. For instance, it is possible to ask, in the terms of this conceit, why leave home at all? Some specific domestic discontent, perhaps, or something vaguer, call it wanderlust. To go where? Somewhere better than, at any rate different from, home, maybe to return in the end, maybe not. What in the philosophy side of the conceit is home? Here, philosophical terminology suggests some candidate answers. Metaphysicians in what is loosely called the analytic tradition speak of ‘naïve metaphysics’, philosophers of mind of ‘folk psychology’, moral philosophers of ‘common-sense

¹ Iris Murdoch, *The Sovereignty of Good* (Routledge, 2001 [1967]), 1.

morality'. It is with moral philosophy that these pages are concerned, and in this area, developing our conceit yields this question: Where should we start in moral philosophy? Answer: We should start from common-sense morality.

Why start there, or as we might say, here? Because there is nowhere else to start, comes the natural answer. To deny this has all the absurdity of the old joke about the Irishman who offers directions while shaking his head and saying 'If I were you I wouldn't start from here.' Why leave at all? Because – and here the apologist for moral philosophy is in a better position than one for (say) metaphysics – there is so much to be dissatisfied with in common-sense morality, the accretions of history, the weight of prejudice, the pervasive irrationality, the apparent inability even of people arguing in good faith to reach agreement. Or, for few philosophers are motivated by reformist zeal, simply out of curiosity, to see what the alternatives to common-sense morality might be, and to see what it might itself look like from the perspective of these alternatives. After all, not every traveller intends to emigrate. But even those who return to their natal villages come back changed in some way, certainly better aware of where their village lies in the wider world and better able to offer directions to other would-be travellers. Most people have never had the opportunity, and many lack the desire, to gain such a perspective. Gilbert Ryle memorably compared them to 'people who know their way about their own parish,

but cannot construct or read a map of it, much less a map of the region or continent in which their parish lies'.²

Ryle, one of the founding figures of the style of modern philosophy with its centre at Oxford, offered a powerful metaphor of the philosopher as *cartographer*. Mapping, as he saw it, was to act as a prophylactic against the errors and omissions of older maps. He recognised that maps, being stylised representations of a more complex reality, could themselves be agents of distortion. His most important work, *The Concept of Mind*, was devoted to a dissection of the Cartesian 'myth', one 'which continues to distort the continental geography of the subject.'³ This cartographical conceit is well worth developing. Consider these elegiac lines from the British literary critic and nature writer Robert Macfarlane:

The commonest map of Britain is the road atlas. Pick one up, and you see the meshwork of motorways and roads which covers the surface of the country. ... Considering the road atlas, an absence also becomes visible. The wild places are no longer marked. The fells, the caves, the tors, the woods, the moors, the river valleys and the marshes have all but disappeared. If they are shown at all, it is as background shadings or generic symbols. ... The land itself ... is indifferent to its pictures and to its picturers. But maps organise information about a landscape in a profoundly influential way. They carry out a triage of its aspects, selecting and ranking those aspects in an order of importance, and so they create forceful biases in the ways a landscape is perceived and treated. ... It can take time and effort to forget the prejudice induced by a powerful map. ... The priorities of the modern road atlas are clear. Drawn by computers from satellite photos, it is a map that speaks of transit and displacement. It encourages us to imagine the land itself only as a context for motorised travel. It warps its readers away from the natural world.⁴

² Gilbert Ryle, *The Concept of Mind*, 60th anniversary edition (Abingdon: Routledge, 2009 [1949]), lx.

³ Ryle, *Concept of Mind*, lx.

⁴ Robert Macfarlane, *The Wild Places* (London: Granta Books, 2008), 8–9.

Consider now this passage from the first page of W. D. Ross's *Foundations of Ethics*

(1939):

I propose to take as my starting-point the existence of what is commonly called the moral consciousness; and by this I mean the existence of a large body of beliefs and convictions to the effect that there are certain kinds of acts that ought to be done and certain kinds of things that ought to be brought into existence.⁵

A little while before him, consider the opening lines of the seventh (and now standard) edition of Henry Sidgwick's *The Methods of Ethics*:

The boundaries of the study called Ethics are various and often vaguely conceived: but they will perhaps be sufficiently defined, at the outset, for the purposes of the present treatise, if a 'Method of Ethics' is explained to mean any rational procedure by which we determine what individual human beings "ought" – or what it is "right" for them – to do, or to seek to realise by voluntary action.⁶

In Sidgwick as in Ross, the question is, roughly, what one ought to do. This sentence, entirely characteristic of Sidgwick's prose, sets the tone for what is to follow in that great work. Sidgwick's prose, here as elsewhere, falls well short of elegance, marred by its subordinate clauses. The qualifications in Sidgwick's prose are like scars of approximation, hinting at complexities left unexplored, even when there is no sense yet of what he might have omitted. His qualifications here – 'sufficiently defined, at the outset, for the purposes of the present treatise' – sound like the

⁵ W. D. Ross, *Foundations of Ethics: The Gifford Lectures Delivered in the University of Aberdeen, 1935-* (Oxford: Oxford University Press, 1939), 1.

⁶ Henry Sidgwick, *The Methods of Ethics*, 7th edition (London: Macmillan, 1907), 1.

confessions of a cartographer embarrassed by how little his map looks like the landscape it is trying to capture, but pleading that (surely) it does the job alright.

No doubt it does do the job, but unlike the jobbing cartographer working to a commission, Sidgwick's constraints were not imposed by someone else. The 'purposes of the present treatise' were *his* purposes, and our question must be why he took these to be his purposes. A few pages later, Sidgwick takes up the question of how much in human life an abstract philosophical treatment of ethics must, of necessity, leave out. We cannot, he says,

... hope to represent in our general reasonings the full complexity of the actual considerations: but we endeavour to approximate to it as closely as possible. It is only so that we really grapple with the question to which mankind generally require an answer: 'What is a man's duty in his present condition?'⁷

But is this the question, *the* question, to which 'mankind generally require an answer'? Or is Sidgwick projecting onto all the human race the peculiar anxieties of the high-minded Victorian? It is certainly a question, and other people than Victorians have had occasion to ask it of themselves. Philosophers before him had made similar observations. Here, for instance, are Kant's remarks in the *Groundwork of the Metaphysics of Morals*:

Thus common human reason is impelled ... from practical grounds themselves, to go outside its sphere and to take a step into the field of practical philosophy, in order to receive information and distinct directions about the source of its principle and its correct

⁷ Sidgwick, *The Methods of Ethics*, 19.

determination in opposition to the maxims based on need and inclination, so that it may escape from its embarrassment concerning the claims of both sides and not run the risk of being deprived, through the ambiguity into which it easily falls, of all genuine ethical principles. Thus even in common practical reason, when it is cultivated, there ensues unnoticed a dialectic, which necessitates it to seek help in philosophy, just as befalls it in its theoretical use; and therefore the first will find no more tranquillity than the other anywhere except in a complete critique of our reason.⁸

The argument here is subtler, but the language remains one of necessity, compulsion, and need. Moral philosophy is in this vision a response to a general human need for guidance and direction. We need maps in morals, as we do in life, to get around – when we do not already know what to do, or find ourselves lost on the motorway. This is true, and these considerations go a long way towards justifying both maps and something like Sidgwick’s conception of the central questions of ethics. The trouble for these conceptions is that we are not always lost on the motorway, and we are not always at a loss for what to do. In such circumstances, the need for maps is somewhat less urgent, and one’s reasons for consulting a map – when there remain any – come from a different source: not directions but (self-)understanding. A good map for these purposes could not responsibly restrict itself to motorways; an ethics of this kind, call it an *impractical* ethics, could not restrict itself to questions about right action.

⁸ Immanuel Kant, *Groundwork for the Metaphysics of Morals*, trans. Allen Wood (Yale University Press, 2002), 21.

In this tradition of ethics, as in this style of cartography, there is this to be said in the case for the defence: maps (or ethical theories like Sidgwick's) do what they set out to do; why cavil at them for failing to do something else? The motorist's map is no good for the rambler or the recreational map-reader, but there are other maps for them with all the rich detail of the landscape beyond the motorway. Indeed there are, but here the analogy takes on a more minatory aspect.

As the road map 'encourages us to imagine the land itself only as a context for motorised travel', so do the ethical treatises quoted above encourage us to imagine ethical experience in terms of questions about our obligations, our reasons for action and so forth. And as the road atlas leaves out 'the fells, the caves, the tors, the woods, the moors, the river valleys and the marshes', so does ethics of this sort leave out, or reduce to generic symbols, much in our experience that has little to do with duty or reasons for action: this is, speaking broadly, the stuff of moral psychology, our habits, dispositions, sentiments – what used to go under the label of 'character'. If one spends long enough on motorways, the rest of the landscape might come to seem, as the history of the industrial revolution has shown, not an independent source of value and interest, but an obstruction and an inconvenience we should be rather better without.

In Sidgwick and Ross, and even in Kant where they are allowed to put up more of a fight, the stuff of moral psychology is ultimately argued or stipulated away as not part of the real questions of ethics as they conceive it. The stipulation is honestly stated in each case, and philosophers are entitled to their stipulations, the more so when they argue for them, as Kant does with some fervour. But the stuff of moral psychology and its study still need somewhere to live, and there is no reason to think their true home is anywhere but in ethics.

Much of what is to follow may be glossed as an extended commentary on what contemporary Anglophone moral philosophy has in common with the British road atlas. I think the narrowing of moral philosophy's focus to questions about action and obligation may be given a historical location – or two – but only as a convenient shorthand for a vastly more complicated historical process (Ryle too conceded that Descartes was not solely responsible for the errors he traced back to him). But this is not the only narrowing that Ross's stipulations involve. The 'moral consciousness' that he takes as his starting point consists in 'a large body of *beliefs and convictions*' about what we ought to do.

It would be a mistake to assume that all of these convictions are true, or even that they are all consistent; still more, to assume that they are all clear. Our object must be to compare them with each other, and to study them in themselves, with a view to seeing which best survive such examination, and which must be rejected either because in themselves they

are ill-grounded, or because they contradict other convictions that are better grounded; and to clear up, so far as we can, ambiguities that lurk in them.⁹

This is, he continues, ‘the time-honoured method of ethics. It was the method of Socrates and of Plato; ... It was the method of Aristotle, and has indeed nowhere been better formulated than it is by him. ... Kant’s method was the same’.¹⁰

Sidgwick seems to have shared both this conception of ethics and this view about its genealogy. In a passage of intellectual autobiography written late in his life, he writes of the moment in his youth when he found himself less than satisfied with the utilitarianism of Mill and Bentham, and therefore, less inclined simply to dismiss their ‘Intuitionist’ opponents as votaries of prejudice and anti-intellectualism:

In this state of mind I had to read Aristotle again; and a light seemed to dawn upon me as to the meaning and drift of his procedure ... What he gave us there [sc. in the *Nicomachean Ethics*] was the Common Sense Morality of Greece, reduced to consistency by careful comparison: given not as something external to him but as what ‘we’—he and others—think, ascertained by reflection. And was not this really the Socratic induction, elicited by interrogation? ... Might I not imitate this: do the same for our morality here and now, in the same manner of impartial reflection on current opinion? ... Indeed ought I not to do this before deciding on the question whether I had or had not a system of moral intuitions? At any rate the result would be useful, whatever conclusion I came to.¹¹

In both these philosophers, we find what I shall call a *doxastic conception* of common-sense morality – roughly, the claim that what is essential to common-sense morality

⁹ Ross, *Foundations of Ethics*, 1.

¹⁰ Ross, *Foundations of Ethics*, 1.

¹¹ *The Methods of Ethics*: xix–xx.

is the set of shared beliefs (convictions, intuitions, judgments) it contains; both trace the conception back to Socrates. Ross's 'moral consciousness' is a set of beliefs; Sidgwick's common-sense morality is 'what "we" ... think', 'current opinion'. Now compare Kant:

... in the moral cognition of common human reason we have attained to its principle, which it obviously does not think abstractly in such a universal form, but actually has always before its eyes and uses as its standard of judgment. It would be easy here to show how, with this compass in its hand, it knows its way around very well in all the cases that come before it, how to distinguish what is good, what is evil, what conforms to duty or is contrary to duty, if, without teaching it the least new thing, one only makes it aware of its own principle, as Socrates did ...¹²

The language is less obviously doxastic (though 'judgment' in early modern philosophy is often roughly coextensive with 'belief' and 'opinion' as used in the subsequent tradition), but the appeal to – and critique of – common-sense morality is again traced back to Socrates. The view persists in an influential modern work such as John Rawls's *A Theory of Justice*, where Rawls declares that 'Moral philosophy is Socratic: we may want to change our present considered judgments once their regulative principles are brought to light'¹³ – Rawls's view of Socrates is, unsurprisingly, Kantian through and through.

¹² Kant, *Groundwork*, 19.

¹³ John Rawls, *A Theory of Justice* (Harvard University Press, 1971), 49. The line does not appear in the revised edition, but there is no obvious reason to think he had changed his view on the question.

The doxastic conception of common-sense morality is – as will generally be agreed even by those who use it – a useful approximation. It can be helpfully contrasted with the more detailed account of common sense it is approximated from, let us call it the *dispositional conception*: common sense not as a set of beliefs but a set of dispositions – entrenched habits of thought, sentiment, action and concept-use that do not easily admit to being reduced without remainder to a set of beliefs. Certainly, the English phrase common sense is often used in the sense of a *skill*, of a certain quotidian but indispensable kind. The anthropologist Clifford Geertz nicely captures the ordinary use of the phrase when he writes:

When we say someone shows common sense we mean to suggest more than that he is just using his eyes and ears, but is, as we say, keeping them open, using them judiciously, intelligently, perceptively, reflectively, or trying to, and that he is capable of coping with everyday problems in an everyday way with some effectiveness. And when we say he lacks common sense we mean not that he ... fails to grasp the fact that rain wets or fire burns, but that he bungles the everyday problems life throws up for him: he leaves his house on a cloudy day without an umbrella; his life is a series of scorplings he should have had the wit not merely to avoid but not to have stirred the flames for in the first place. The opposite of someone who is able to apprehend the sheer actualities of experience is ... a defective; the opposite of someone who is able to come to sensible conclusions on the basis of them is a fool. And this last has less to do with intellect, narrowly defined, than we generally imagine.¹⁴

A description of common-sense morality might restrict itself to its doxastic components, as the map of a landscape might restrict itself to paved motorways. But we should be clear in our minds that this involves a choice, and an omission, with the power to shape perceptions of the phenomena.

¹⁴ Clifford Geertz, “Common Sense as a Cultural System,” in *Local Knowledge* (New York: Basic Books, 1983), 76.

If a landscape has motorways in it, it is because they have been put there.

Sometimes they are built on older roads built by (say) Romans from more than a thousand years ago, themselves the formalisation of informal paths formed by walkers repeatedly treading on certain patches of grass. In that sense, they will often reflect some older, more basic, feature of the landscape – a ridge, a plateau, or a slope. But it is always a question when this is the case. Similarly, the particular dispositions, of forming convictions or judgments about the rightness and wrongness of actions, are often already there in common-sense morality. But it is an open question whether these dispositions, once they have been overlaid with their doxastic interpretations, retain their original shape, and indeed, to what extent the psychological landscape is improved by the overlay. I am suggesting, in other words, that we should not without argument assume common-sense morality to have the structure attributed to it by philosophers who assume the doxastic conception, *a fortiori* a doxastic conception committed to the centrality of a particular set of beliefs.

This dissertation was conceived, and began life, as a commentary on themes in Henry Sidgwick's *The Methods of Ethics* with relevance for questions in contemporary moral philosophy. As time went on, I came to the view that there was in fact a single

theme in Sidgwick more important than, and able to subsume, all others. This was the aspiration that gave his book its structure: the idea that a careful and rigorous examination of ‘the morality of common sense’ that looked beyond its superficial diversity might well find a deeper unity in basic, self-evident axioms of practical reason that might command universal rational assent. In aiming to do this, Sidgwick effectively set the terms of virtually every important debate in twentieth-century Anglo-American moral philosophy. His influence is amply evident in the writings of John Rawls, Derek Parfit and Peter Singer, all of whom speak of Sidgwick with great respect.¹⁵

Part of Sidgwick’s influence can be observed simply in the ease with which his arguments can be assimilated into contemporary philosophy; this is certainly not the case with a work such as F. H. Bradley’s *Ethical Studies*, published not long after *The Methods of Ethics*, but well-armoured by its style against any attempt to pressgang its arguments into some contemporary debate. Modern Anglophone moral philosophy is Sidgwick’s child, if it is anyone’s, and this influence can be seen in the contemporary ubiquity of a set of philosophical methods and styles that

¹⁵ See the editor’s forward to John Rawls, *Lectures on the History of Political Philosophy*, ed. Samuel Richard Freeman (Harvard University Press, 2009), xi for a brief account of Rawls’s regard for Sidgwick; the Preface to Derek Parfit, *On What Matters, Volume I* (Oxford: Oxford University Press, 2011), 46–56; Peter Singer, “Sidgwick and Reflective Equilibrium,” *The Monist* 58, no. 3 (1974): 490–517 and more recently the co-authored Katarzyna de Lazari-Radek and Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford: Oxford University Press, 2014).

Sidgwick did so much to develop and make rigorous, among them the attempt to elicit from ordinary moral thought, discourse and practice a set of beliefs, or considered judgments, or intuitions, that are then put to the test, modified in light of the demands of other such beliefs, judgments and intuitions on the one hand, and the demands of more general principles of theoretical goodness on the other.

Rawls's label for this general style of moral philosophy, 'reflective equilibrium', is the closest thing that contemporary ethics has to a standard method.

In addition to this idea of method, we find in Sidgwick a highly distinctive style of thought and writing, one characterised by a relentless aspiration to detachment from his own beliefs and sentiments, and a tendency to push (what we should now call) the principle of interpretative charity to almost fanatical lengths. Above all, we find in Sidgwick an attitude to common-sense morality just as characteristic of modern moral philosophy. This attitude stems from the idea that there is nowhere else to start in ethics but in the beliefs taken to be constitutive of common-sense morality, but that it must always be an open question whether philosophy will leave common sense where it was. As Derek Parfit briskly puts it, 'Philosophers should not only interpret our beliefs; when they are false, they should change them'.¹⁶

¹⁶ Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), x.

A philosophical tradition could do rather worse than to have Sidgwick for a progenitor. Sidgwick, self-effacing as always, did not make any grand claims for his own originality, and seems not to have sought influence or celebrity; indeed, his views committed him to reticence about his heterodox ethical and religious (that is to say, irreligious) convictions. He saw himself as belonging to a certain intellectual tradition, and wrote with unfeigned respect of his indebtedness to his predecessors in that tradition, one that – as we have seen – he supposed to go back to Socrates. If his own part in that tradition, both as contributor and historian, have long been underrated, they are in good part consequences of that reticence.¹⁷

One example of Sidgwick failing to receive his due is the phrase, ‘common-sense morality’, itself. So common are references to it in modern moral philosophy that it is possible to doubt that there was ever a point at which the phrase needed to be introduced into philosophical discourse.¹⁸ There is one sense, as we have seen, in which the doubt is justified: as long as there have been philosophers, there have been non-philosophers (variously, and ambivalently, ‘the naïve’, ‘the vulgar’, ‘the folk’), and they have had their ways of carrying on, and their ways of talking about

¹⁷ These themes are explored in great depth in Bart Schultz, *Henry Sidgwick – Eye of the Universe: An Intellectual Biography* (Cambridge: Cambridge University Press, 2004).

¹⁸ The best recent treatment of the concept of common sense in history, with a special focus on the early modern period, is Sophia Rosenfeld, *Common Sense: A Political History* (Cambridge, Mass: Harvard University Press, 2011).

their ways of carrying on. Equally, there was much talk of a ‘philosophy of common sense’ among philosophers of the so-called Scottish Enlightenment. Writers before Sidgwick, such as his *bête noire* William Whewell, had employed the general idea in moral philosophy, but it is Sidgwick who gives it the pithy name we continue to use:

By philosophic minds [writes Sidgwick] ... the ‘Morality of Common Sense’ (as I have ventured to call it), even when made as precise and orderly as possible, is often found unsatisfactory as a system, although they have no disposition to question its general authority. It is found difficult to accept as scientific first principles the moral generalities that we obtain by reflection on the ordinary thought of mankind, even though we share this thought. ... without being disposed to deny that conduct commonly judged to be right is so, we may yet require some deeper explanation *why* it is so. From this demand springs a ... phase of Intuitionism, which, while accepting the morality of common sense as in the main sound, still attempts to find for it a philosophical basis which it does not itself offer ...¹⁹

Note how, in the course of a single paragraph, the ‘Morality of Common Sense’ – ‘the moral generalities that we obtain by reflection on the ordinary thought of mankind’ – has lost both the inverted commas and the capital letters, assimilated thoroughly into the parlance of philosophy.²⁰ Book III of *The Methods of Ethics* is where Sidgwick spends longest giving an account of the generalities implicit, or sometimes explicit, in common-sense morality. Time and again, he finds these beliefs wanting, but is at pains to insist that he has done his best to represent common-sense morality in its best light. If it continues to seem inadequate, he

¹⁹ Sidgwick, *The Methods of Ethics*, 101–2; emphases in original.

²⁰ More recently, it is very likely its use in the early writings of Derek Parfit (e.g. Derek Parfit, “Is Common-Sense Morality Self-Defeating?,” *The Journal of Philosophy* 76, no. 10 (October 1, 1979): 533–4) that brought the phrase to new prominence.

implies, it is because it is. And if Sidgwick's manner is inquisitorial, we should remember that he too stands charged with the rest of us:

The morality that I examine in Book iii. is my own morality as much as it is any man's: it is, as I say, the 'Morality of Common Sense,' which I only attempt to represent in so far as I share it; I only place myself outside it either (1) temporarily, for the purpose of impartial criticism, or (2) in so far as I am forced beyond it by a practical consciousness of its incompleteness.²¹

Sidgwick does not think this style of philosophy an innovation of his. His own literary style, detached and largely affectless, is Aristotelian; so is his method, of giving us 'what "we" ... think, ascertained by reflection', combining a stern impartiality with the guarded solidarity suggested by his pronoun. This is opposed to Aristotle's most immediate inspiration, those among the writings of Plato that are commonly taken to present the practice of the historical Socrates. Where Aristotle 'ascertained' common-sense morality 'by reflection', in Plato, it is 'elicited by interrogation': or so Sidgwick says.

This characterisation of Plato and Aristotle marks a methodological contrast between them, but not one to which Sidgwick attributes much significance.

Aristotle, he says,

retained from Plato's teaching the original Socratic method of induction from and verification by common opinion. Indeed, the turns and windings of his exposition are best understood if we consider his literary manner as a kind of Socratic dialogue formalised and

²¹ From the preface to the 2nd edition: Sidgwick, *The Methods of Ethics*, xii.

reduced to a monologue—transferred, we may say, from the marketplace to the lecture-room.²²

When Plato's Socrates wished to have the content of common-sense morality explicated, he asked people; Aristotle, by contrast, seems to have thought he could do it by himself, and Sidgwick follows him in this. Neither seems to have thought their armchair anthropologies raised any special difficulties. Sidgwick for one is evidently anxious to fend off accusations that his picture of common-sense morality is in any way idiosyncratic: 'I wish it to be particularly observed, that I have in no case introduced my own views, in so far as I am conscious of their being at all peculiar to myself'.²³

One might be inclined to be suspicious of Sidgwick, as of any unelected spokesman, however insistent he be of his integrity. But we should be clear that it is an open question whether the Aristotelian-Sidgwickian position on the matter, that 'reflection' is sufficient to distinguish idiosyncratic opinions from those shared with other people, is in fact inferior to the quotidian empiricism of Socrates who took the trouble actually to ask people what they thought. In the first of the two long essays that constitute this thesis, I look to Plato's dialogues to discuss some of the deep problems raised even by so innocent a method as Socrates'. In brief, the problem is

²² Henry Sidgwick, *Outlines of the History of Ethics for English Readers* (London : Macmillan, 1886), 55–6.

²³ Henry Sidgwick, *The Methods of Ethics*, 1st edition (Cambridge University Press, 1874), 317.

this: given Socrates' well-known scepticism about people's claims to self-knowledge, why should they be treated as authoritative spokespersons about their beliefs?

Plato's Socrates, as we well know, engaged in a series of conversations with ordinary interlocutors about – among other things – such everyday notions as those of courage, beauty and justice. He sought definitions from his interlocutors, in part because they claimed, usually with unfounded confidence, to know about such matters. When they replied to Socrates' questions, it is usually hinted that they were not giving voice to a merely idiosyncratic opinion; they became, often without knowing it, spokesmen for the conventional beliefs, attitudes and practices of their society. The trials those beliefs and attitudes underwent at Socrates' hands were more than *ad hominem* attacks on their possessor's cogency, though they certainly were at least that; they were instances of what we may well call *critique*.

It is usually claimed that the Socrates in Plato's dialogues drew the tools of his critique not from some alternative heterodox morality of his own devising but from *within* the morality that was the object of critique; his methods were, in the jargon, 'peirastic'. Much progress could be made simply by exposing the internal inconsistencies and ambiguities of the beliefs and attitudes that constituted that morality. It is a conspicuous characteristic of the dialogues of Plato's usually

classified as 'early' and 'Socratic' – that is to say, written early in Plato's career and perhaps for that reason likelier to reflect an attempt at representing the practice of the historical Socrates – that they lack the methodological self-consciousness of his 'late' style, a self-consciousness and reflexivity quite overwhelming in such dialogues as the *Sophist* and *Statesman* whose subject matter is their own methods. The figure of Socrates in the early dialogues does seem to have a method, after a fashion, but the more assiduously modern scholarship has tried to reduce that method to a set of formal criteria, the less they have been able to capture the freewheeling, improvisatory quality of these dialogues.

In these dialogues, Socrates asks questions and exposes inconsistencies in his interlocutors' answers, but that is not all he does. He also tells jokes and stories to his interlocutors, makes edifying remarks to them, teases them, even flirts with them, sometimes outrageously. Certainly he elicits their beliefs from them by interrogation, though it is a serious question just how what is involved in something's being a belief of theirs, but that is not all he elicits. One of Plato's many literary gifts was one for characterisation, and his Socrates is a vivid creation, as are many of his interlocutors. Their beliefs are certainly part of their character, but in Plato's presentation, they are embedded in a wider range of dispositions of behaviour, thought and sentiment. In this sense, Plato's form is superbly well suited

to the substance of his moral philosophy, his dialogues able not just to announce the results of an inquiry into moral psychology but to dramatise that inquiry itself by showing us moral psychology embodied in individual character(s).

Aristotle's works, which Sidgwick insightfully saw 'as a kind of Socratic dialogue formalised and reduced to a monologue', do not present a moral philosophy radically at odds with Plato's. Certainly there are differences in the underlying metaphysical picture, but Aristotle shares with Plato a moral philosophy deeply entwined with the fundamental categories of moral psychology – the different forms of human motivation, rational and non-rational – and presented as part of a picture that contains also a theory of moral education and a vision of politics. Yet, the form of Aristotle's surviving works (his dialogues are lost), marked by the air of the lecture-room, do not allow the happy congruity between form and substance we find in Plato. Aristotle was as concerned as Plato was with human character and the role that non-rational motivations play in the ethical life. But his descriptions of method, however continuous with those of Plato's Socrates, do leave something out.

An orthodox understanding of Aristotle's most distinctive method of doing philosophy finds a succinct statement of that method's animating principle in Book VII of his *Nicomachean Ethics*:

As in the other cases, we must set down the phenomena, and then, having first gone through the puzzles, prove in this way (if possible) all the reputable opinions (*endoxa*) about these conditions, or else (if that can't be done) the greatest number of them and the most important. For if the difficulties are resolved, and the reputable opinions are left, that will be proof enough. (1145b2–7)²⁴

‘Endoxic’ inquiry, as we might call it, may be represented schematically as a three-stage procedure. First, one gathers a set of *endoxa*, or reputable opinions, on some subject. This includes beliefs held in explicit form by all, or nearly, all people, the implicit presuppositions of their behaviour, the commitments behind their language-use and discourse, and even the minority opinions of people with some claim to be taken seriously. Second, one looks among the *endoxa* for infelicities. These may include inconsistencies or tensions between two or more reputable opinions, or instances of vagueness, ambiguity, or simple implausibility. Finally, one attempts to resolve the infelicities discovered in the previous stage by restating the *endoxa* so that they are free of ambiguity or vagueness, and by abandoning some of them so as to leave behind a ‘maximal consistent subset’ of the most important of the original *endoxa*.²⁵

What is missing in Aristotle that is present in Plato is some visible sense of what is involved in the first step in this process: the gathering of the *endoxa*. In Aristotle,

²⁴ The translation is by Jonathan Barnes, quoted in Jonathan Barnes, “Aristotle and the Methods of Ethics,” in *Method and Metaphysics: Essays in Ancient Philosophy I* (Oxford: Oxford University Press, 2011), 174.

²⁵ The quoted phrase is from Barnes, “Aristotle and the Methods”, 177.

this is always a *fait accompli* – and why not? Surely the task poses no special challenge to a moderately self-conscious, intelligent mind with some experience of life – or so he might have thought. But in Plato, the gathering of opinions, or something like it, is itself part of the business of philosophy. The dialogues themselves do not of course *do* this so much as present it being done; they are dialogues, literary artefacts showing a high degree of authorial design, and not a sociologist's questionnaires. Yet, part of the dialogues' evident authorial design involves the dramatisation of the processes of philosophising, even if Socrates' philosophising stands at a remove from Plato's own.

In moving from Aristotle to Sidgwick, one leaves out much that is important and interesting in the history of philosophy, but it is how Sidgwick in some moods constructs his own intellectual genealogy. As one footnote has it, 'through a large part of the present work the influence of Plato and Aristotle on my treatment of this subject has been greater than that of any modern writer'.²⁶ It is not that he did not take the other figures of the western canon seriously: he discusses Hume and Butler with some respect, and the influence of Kantian notions of universalizability on his moral philosophy are richly evident. Still, on matters of method in particular, it is to

²⁶ Sidgwick, *The Methods of Ethics*, 375 n. 2.

the two Greeks that he constantly defers, more often than to any recent predecessor or contemporary.

However, for all that he claimed for his philosophy a classical pedigree, Sidgwick's moral philosophy was in several ways deeply unclassical. In part, this was an aspect of his self-conscious modernity. His idea of a 'method of ethics', as he defines it in the opening sentences of his great work quoted earlier, was 'any rational procedure by which we determine what individual human beings "ought" – or what it is "right" for them – to do, or to seek to realise by voluntary action.'²⁷ This was, as he would later note, a question conceived in distinctively modern, rather than ancient, terms:

Their [sc. the ancients'] speculations can scarcely be understood by us unless with a certain effort we throw the quasi-jural notions of modern ethics aside, and ask (as they did) not "What is Duty and what is its ground?" but ... 'What is the relation of the kind of Good we call Virtue, the qualities of conduct and character which men commend and admire, to other good things?'"²⁸

His *Outlines of the History of Ethics for English Readers*, a short book revised from his article on the history of ethics for the *Encyclopaedia Britannica*, has a sophisticated account of the part played in the history of late antiquity and mediaeval Europe by

²⁷ Sidgwick, *The Methods of Ethics*, 1.

²⁸ Sidgwick, *The Methods of Ethics*, 102. For an argument against this way of distinguishing ancient from modern ethics, see Roger Crisp, *Reasons and the Good* (Oxford: Oxford University Press, 2006), 27–35, who argues that all the materials for a 'quasi-jural' picture of ethics are to be found in Plato and Aristotle. It is not important for my purposes whether Sidgwick's contrast was well-drawn; I quote the passage as evidence of his self-understanding.

ideas of morality with their roots in Judaism. *The Methods of Ethics*, though its classical debts are large and numerous, is not the work of a reactionary Philhellene. It is too evidently dissatisfied with the apparent triviality of ancient answers to the question he thought it the aim of moral philosophy to address:

... we ask how we are to ascertain the kind of conduct which is properly to be called Virtuous, it does not seem that Plato can tell us more of each virtue in turn than that it consists in (1) the knowledge of what is Good in certain circumstances and relations, and (2) such a harmony of the different elements of man's appetitive nature, that their resultant impulse may be always in accordance with this knowledge. But it is just this knowledge (or at least its principles and method) that we are expecting him to give us: and to explain to us instead the different exigencies under which we need it, in no way satisfies our expectation. Nor, again, does Aristotle bring us much nearer such knowledge by telling us that the Good in conduct is to be found somewhere between different kinds of Bad. This at best only indicates the *whereabouts* of Virtue: it does not give us a method for finding it.²⁹

Even in Aristotle's more systematic presentation of his ethics, much work is left to that demanding virtue, 'phronesis' – prudence, wisdom, or judgment – which does not, in Aristotle's discussions of it, seem to be amenable to statement in the form of a simple rule. There is little to be said in defence of Plato and Aristotle against the charge of triviality once Sidgwick's demands of a method of ethics are accepted as legitimate.

The question of whether they *are* legitimate has been raised off and on in modern ethics, like a discordant note sounding every so often in an otherwise harmonious orchestration – by, among others, such figures as Iris Murdoch, Elizabeth Anscombe

²⁹ Sidgwick, *The Methods of Ethics*, 375–6.

and Bernard Williams, on whom I shall have more to say. For the most part, however, Sidgwick's demands have been accepted, and indeed, taken to be constitutive of the discipline of ethics itself. It is certainly the standard way of teaching ethics to undergraduates in Anglo-American philosophy departments, leavened only slightly by the distinction owed to Sidgwick's most influential successor, G. E. Moore, between ethics and *metaethics*. Sidgwick, like Mill and numerous others before him, did not trouble himself greatly with the distinction. In his case, it was in part because he took a minimal set of metaethical views for granted, views that might be characterised in anachronistic terms as 'realist', 'cognitivist' and (with less textual warrant) 'non-naturalist' in their general orientation. Of course, not even Sidgwick, with all his deserved reputation for inventing wheels his successors would nevertheless expend much sweat on reinventing, gives us a sufficiently determinate basis for attributing to his views much more precision than that. Nevertheless, as far as 'first-order' ethics is concerned, Sidgwick took his central task – roughly, finding a rational procedure to determine what one ought to do – to be one that could be accomplished, and most philosophers in the Anglophone tradition have followed him in thinking this possible and of central importance. There remains some place for moral psychology in ethics as thus construed, but the deep connection between philosophy and psychology in Plato and Aristotle has been effectively severed. Sidgwick's picture of

ethics gives it an autonomy from psychology that it did not have in the ancient world.

The 'psychological turn' in Plato, as we might well term it, comes from the contexts in human life that his dialogues represent as motivating ethical reflection. These are, centrally, the needs of education and those of politics, in particular, of political stability; the *Meno*, *Protagoras* and *Republic* exemplify these preoccupations in their purest form. It is natural that an adequate answer to these demands would call for a deep engagement in facts about the human dispositions generally regarded together under the traditional label of 'character'. (Indeed the word 'ethics' is etymologically connected to the Greek for character.) This remains true in Aristotle: his works lack the dramatic contexts of Platonic dialogues that make the link apparent, but his work is too full of asides on pedagogy and politics to be regarded as an abandonment of the Platonic position on the matter. If Plato and Aristotle have – to use a deliberate anachronism – an ethical *theory*, then it is a theory whose structure is responsive to the structure of the psyche. In Sidgwick, however, the motivations for ethical reflection seem rooted in the need for guidance in practical deliberation and his ethical reflection delivers strictures with a correspondingly austere, and largely action-focused, structure. The most general answer to the ethicist's basic question in the post-Sidgwick tradition takes the form 'An action is

right iff it...'; when set aside Aristotle's account of virtue as a disposition consisting in a mean defined by reference to the reason with reference to which the prudent person would define it, this tradition embodies the modern hope that deference to persons, even as heuristics, indeed deference to anything other than rationality itself, can finally be done away with.

Sidgwick thought his conception of ethics no external imposition but one characteristic of ordinary moral life at its best (recall Kant's remark on how the cultivation of ordinary practical reason produces a dialectic that leads it to philosophy). Indeed, when he looks to describe the content of common-sense morality, what he hopes to uncover is the rules and 'formulae' in which it is expressed. The idea is to try to express the dispositions constitutive of common-sense morality in well-articulated propositions of the sort demanded by (what I have called) the doxastic conception of common sense. '[S]uch formulae are not difficult to find', says Sidgwick breezily. '[I]t only requires a little reflection and observation of men's moral discourse to make a collection of such general rules, as to the validity of which there would be apparent agreement at least among moral persons of our own age and civilisation, and which would cover with approximate completeness the whole of human conduct'.³⁰

³⁰ Sidgwick, *The Methods of Ethics*, 214–5.

We might wonder about the extent and significance of his restriction of his sample to ‘moral persons of our own age and civilisation’, but we shall let that pass. Let us dwell instead on Sidgwick’s conceiving of common sense as essentially a set of principles, indeed a set of *general* principles. Sidgwick’s reasons for taking this view of common-sense morality are, briefly, that ‘general rules’ are what we turn to when we are confronted with difficult cases, and moreover, that while particular judgments are liable to be corrupted by ‘the complexity of circumstances’, ‘our personal interests or habitual sympathies’, the general rules we endorse are at least less liable to be so corrupted.³¹

How much do these considerations really show? The second consideration shows little by itself: ethical judgments made at a high degree of generality stand in danger of failing to do justice to the basic ethical dispositions they are supposed to be expressing just as (say) general grammatical or orthographical claims fail to do justice to the linguistic competence of a native speaker. The first argument seems to rest on this premise: when the ordinary person, confronted with a situation where she must *think* about what she is to do, the principles to which she adverts in her thinking on the question will be revelatory of her morality; when such principles

³¹ Sidgwick, *The Methods of Ethics*, 214.

are found to be widely shared, they may be treated as constitutive of common-sense morality. But it is possible that the case Sidgwick describes, in which the moral agent needs to think about what the right thing to do is, is in a quite special situation. Such a situation ought not to be treated without argument as either paradigmatic of the role of morality in everyday life or revelatory of one's moral principles.

Should I keep this promise, I wonder. Of course I should, after all I promised, says common-sense morality. Should I break this promise or that one, I wonder, given that I cannot keep both. Common-sense morality goes as far as to show me that I am indeed in a fix: what ought I to do? But once I start asking myself, say, which promise it would do the greater harm to break, or to whom to I owe the greater duty, or who would resent it less, we have – I propose – stepped out of the province of common-sense morality proper and entered into the sphere of moral reflection, practical judgment, ethical improvisation. No doubt the principles, or quasi-principles, to which we sometimes defer in such reflection are revelatory of something. But it is at least as plausible that they reflect simply a practical need to settle the question in some non-arbitrary way. In referring the matter to a principle, we have taken one step away from common sense. Insofar as common-

sense morality does not point us in any obvious direction, it seriously underdetermines the conclusion of the moral deliberation.

This is not a critique of a rule-based morality, or of the importance of arbitrating principles in moral reasoning. It is merely an argument against – firstly – making it a requirement on something's being part of the morality of common sense that it be expressible in the form of a general rule, and against – secondly – supposing that the general rules putatively constitutive of common-sense morality are revealed in moments of moral reasoning about difficult situations. Sidgwick does not say enough to disarm the thought that what seem to be flaws in common-sense morality (vagueness, inconclusiveness) are less damning than they seem: they are equally evidence that common-sense morality has in it space for judgment. If common sense has a principle on the matter, it is the (sensible) position that such cases cannot be decided at the level of generality at which Sidgwick's examples are described. If Sidgwick thinks they can be, he needs a stronger argument than the fact that utilitarianism *can* in fact deal with them at this level of generality, because whether this should be taken to be a theoretical virtue or not is itself a matter for argument.³²

³² A similar range of arguments against Sidgwick is elaborated in Marcus G. Singer, "Sidgwick and Nineteenth-Century Ethical Thought," in *Essays on Henry Sidgwick*, ed. Bart Schultz (Cambridge: Cambridge University Press, 1992), 76-83.

On an alternative way of thinking about the matter, the proper heuristic to reveal the content of common-sense morality is not those situations in which we must think about what to do; it is those situations in which it is perfectly obvious what to do. And for all others, common-sense morality steps back gracefully to make way for – what? – wisdom? judgment? theory? So Sidgwick might be right ‘not [to] accept Butler’s view as to the sufficiency of a plain man’s conscience: for it appeared to me that plain men agreed rather verbally than really’. But if we replace the ‘plain man’ of Butler’s imaginings with the *wise* one, then the claim reads more plausibly, though it does, of course, leave all the important questions open.

In Plato, the structure of the ethical theory is well suited to both its subject matter and its literary expression in the complexity, richness and detail of its psychological descriptions. In Aristotle, the structure of the theory remains well suited to its subject matter though the relative dryness of the literary presentation makes it less easy to see that this is so. In Sidgwick, we find again a harmony of theoretical structure, subject matter and literary style, but this has been achieved by a reconstitution of the subject matter itself: not the relatively open-ended question one finds in Plato’s Socrates – ‘how should one live?’ – with the interrogative

pronoun hinting strongly at an answer cast in psychological terms, but the more specific question of what one should do and why. The questions are not equivalent, but it is possible that everything of significance in the Socratic question can be captured in Sidgwick's. But if this is true, it must depend on an unargued view about the relationship between action and character and the importance for the evaluation of an action of facts about the dispositions that originate the action.

None of this has shown, or tried to show, that Sidgwick's question is a bad one for ethics to focus on. It is, rather, an attempt to place Sidgwick's question into the broader geography of human concerns, to which the ancient philosophers were sensitive, and to which a heterodox strain in ethics after Sidgwick have striven to return. The two essays that comprise this dissertation are extended commentaries on the historically Janus-faced quality in Sidgwick: the way in which he looks *backward*, to the origins of western moral philosophy among the ancient Greeks, and the way in which his conception of ethics – Socratic practice reinterpreted to meet the needs of philosophy in modernity – points *forward* to the tradition of Anglophone academic ethics he did so much to create. It can seriously be argued that Anglophone ethics since 1900 (the year of Sidgwick's death) is the house that Sidgwick built, but to defend this claim about diachronic influence needs a different kind of history than mine, one more willing than mine has been to rootle around in

archives, memoirs, university syllabi and questions in old examination papers. Even a history in a more ‘analytic’ mode, that is to say one concerned only with decontextualised argumentation in some set of putatively canonical texts, would need to say more than I do about how ideas from Sidgwick figure in Moore, Prichard, Ewing, Joseph, Ayer, Stevenson, Broad and Ross – to name a handful of philosophers influential in the first half of the twentieth century.³³ I pass over these figures in favour of a more recent tradition, one centred around Oxford in the years after the end of the second world war.

Of these, I am most interested in Bernard Williams, a philosopher who occupied Sidgwick’s Chair at Cambridge and devoted much of his career to arguing for the futility of what Sidgwick took to be the central aims of moral philosophy. Here my aim is to reconstruct, as clearly as I can, Williams’s much misunderstood critique of moral philosophy in Sidgwick’s style, and consequently, to clarify just to what extent Williams’s alternative involved a reinstatement of the common-sense morality to which Sidgwick, and Socrates, sought an alternative. In doing this, I have been more mindful of the need to bring out the significance of Williams’s corpus not only as part of the *longue durée* history of western moral philosophy from

³³ Some of these (unjustly forgotten) philosophers are the subject of essays in Thomas Hurka, ed., *Underivative Duty: British Moral Philosophers From Sidgwick to Ewing* (Oxford: Oxford University Press, 2011).

Socrates onward, but also within its more local, mid-century British (and often, even more parochially, Oxonian) context. This context includes both the briefly regnant tradition of linguistic philosophy exemplified by J. L. Austin and Gilbert Ryle, and in a different way, by R. M. Hare, as well as the most powerful critics of the regnant tradition as it ramified in ethics, Elizabeth Anscombe and Iris Murdoch. The parochialism of this focus is justified, if a justification is demanded, by the simple fact that mid-century Oxford was, as far as philosophy was concerned, an almost uniquely powerful parish.

The second essay in this thesis is an examination of how Williams (and in different ways, Murdoch, Anscombe and others), put together an alternative interpretation of just what an ethics that sees itself in a Socratic lineage can be. The moral philosophy of Williams, as I reconstruct it, retains the critical spirit that Sidgwick, among other Victorians, saw in Socratic practice, but directs it at common-sense morality conceived more capaciously as consisting in psychological dispositions that are not (just) the disposition to follow a rule or apply a principle. Once moral psychology is reinstated to the position of importance it had for Plato and Aristotle, the critical, revisionary aims of moral philosophy must involve more than replacing false ethical beliefs with true ones. The task would be to give a full account of the structure of common-sense morality – that is to say, of the variegated dispositions

that comprise it – and its place in a wider schema that includes other branches of philosophy, in particular, metaphysics, epistemology, the philosophy of mind, language and science, and other humanistic disciplines, in particular history and anthropology.

Williams's work, I claim, is propaedeutic, hinting at a philosophical enterprise he did not do much more than clear the way for. The central idea in his corpus is roughly the one with which I began. To wit, that much moral philosophy is in the business of mapping our moral consciousness, and that previous philosophical maps – for reasons that have to do both with the costs of abstraction and the historical and political priorities built into the activity – have distorted that consciousness in ways from which a better philosophy might be able to free us. It is tempting to think of this conception of philosophy as conservative, insofar as it lays part of the blame for the distortion on philosophy itself and seems to promise a return to some pristine, pre-philosophical innocence, but this would be a serious misreading. Kant put the case against innocence well when he wrote:

There is something splendid about innocence, but it is in turn very bad that it cannot be protected very well and is easily seduced. On this account even wisdom—which consists more in deeds and omissions than in knowledge—also needs science, not in order to learn from it but in order to provide entry and durability for its precepts.³⁴

³⁴ Kant, *Groundwork*, 20.

Williams might well have agreed, and added that one of the things from which innocence stood in need of protection was precisely such a philosophy as Kant's, which, as he saw it, distorts the very moral consciousness it claims to defend. His writing is animated by the same spirit that animated Socrates and Plato, Kant and Sidgwick, the spirit of critique. If he gives the impression of unremitting hostility to their philosophies, it is because he thought they did not go far enough.

A final word: no dissertation can be anything but a disappointment to its author, particularly when the reality of the final draft is set against the optimism of one's original plans. The writing of a dissertation is an excellent cure for the Cartesian hubris of the prospectus. In time, one comes to accept with a better grace just how much of an achievement it is just to 'clear up the ideas of one or two people, a little', as Sidgwick once put it.³⁵ I am better able to acknowledge now that Sidgwick's world-historical importance, as part of what a generation of French historians termed '*l'histoire des mentalités*', simply pales in comparison to that of (for instance) Kant and Nietzsche. Yet, we are, all of us, compelled to start our philosophising from where we are. Sidgwick represents the contemporary Anglophone ethical tradition in which I have received virtually all of my philosophical education. In

³⁵ Arthur Sidgwick and Eleanor Mildred Sidgwick, *Henry Sidgwick: A Memoir* (London: Macmillan, 1906), 284.

giving this work the subtitle 'Problems from Sidgwick', I have tried to suggest that it is this tradition itself that is the real subject of my thesis.

Part I

Must we say what we think? Socrates, self-knowledge, and common-sense morality

What [Aristotle] gave us there was the Common Sense Morality of Greece, reduced to consistency by careful comparison: given not as something external to him but as what 'we'—he and others—think, ascertained by reflection. And was not this really the Socratic induction, elicited by interrogation?

Henry Sidgwick³⁶

There is something in critique which is akin to virtue.

Michel Foucault³⁷

1. Introduction

Consider the old examination chestnut: 'Is Plato a radical or a conservative?' The question would not be worth asking were there not something about him that makes it hard to say. His dialogues, where he famously never appears *in propria persona*, are full of paradoxes and revisionary suggestions about everything from morality to metaphysics. Despite what *hoi polloi* ('the many') say, his principal

³⁶ Henry Sidgwick, *The Methods of Ethics* (London: Macmillan, 1907), xix–xx.

³⁷ Michel Foucault, *The Politics of Truth*, ed. Sylvère Lotringer, trans. Lysa Hochroth and Catherine Porter (Los Angeles, CA: Semiotext, 2007), 43.

speaker (usually Socrates) urges that no one does wrong knowingly, no one desires what is bad for them, perception is an unreliable route to knowledge, and middle-sized dry goods don't strictly speaking exist. Yet, his Socrates is decidedly non-radical about a good many things. His standards of knowledge are high, yet he remains invested in philosophical inquiry; his idea of virtue is demanding, he still thinks it worth cultivating; his views about the gods are not exactly orthodox, yet he is given to ostentatious displays of piety. His conservatism comes out particularly sharply when he is contrasted with another anti-commonsensical party.

Time after time in Plato's dialogues, he confronts that familiar creature of the 5th century BC, (let us call him) the sceptic, who never found a commonsensical notion he did not want to undermine. Sometimes, he appears as an impish mountebank, trading on ambiguities in the ordinary uses of words. Sometimes, he appears as a thoughtful proponent of some form of relativism. In his most unsettling avatar, he is an immoralist, rejecting conventional ideas of virtue and vice for a darker picture of conventional values and human motivation. Plato's Socrates shares not at all in this worldview, though he is at one with its hostility to a kind of complacency.

The scholarly consensus about Plato's temperament seems to be a sort of *via media* between the extreme choices in our original question. Is Plato a radical or a conservative? Maybe he is neither, and both. According to one influential *rapprochement* of the two positions, Plato's achievement lies in his going beyond the terms of what was thought to be a choice between antithetical options. He wants neither complacency nor immoralism; he seems to think the traditional faith in an objective basis for ethics (and much else) can be vindicated against the sceptic, but in doing so, will defend the objective truth of many claims that will strike no one as commonsensical. Ordinary practices do not make it through a Platonic dialogue unscathed; indeed, it is possible that Plato thought it essential to the tenability of some ordinary practice in the face of sceptical doubts that it be given a basis in something not itself part of the ordinary practice. In that, he makes an important concession to the radical impulse. The obverse is also true: it often turns out that the practices reconstituted in light of the sceptic's doubts will be recognisably continuous with those under challenge.³⁸

³⁸ This sort of position is defended in Julia Annas, *An Introduction to Plato's Republic* (New York: Oxford University Press, 1981): '[Plato] is the sort of ally that conservatives least want. If he reinstates ordinary moral views, it is on a new basis which is remote from anything that the ordinary person would dream of' (9); and 'Plato is making the brave attempt both to establish that there are objective moral truths and to undertake a considerable overhaul of what those objective moral truths are' (10). A slightly different but compatible claim is made in Terence Irwin, "Common Sense and Socratic Method," in *Method in Ancient Philosophy*, ed. Jyl Gertzler (Oxford: Oxford University Press, 2001), 63.

How Socrates strikes us in Plato's dialogues can be a function of whom he is talking to. To put it crudely: in conversation with the immoralist he can seem deeply conservative, in conversation with the unreflective or conformist he can seem radical. However, his approaches to both sorts of interlocutor have this in common: seldom does he simply assert his views to them. His methods are essentially *peirastic*; they are responsive to what his interlocutors say and feel and do, and this fact is reflected in Plato's most distinctive formal choice, the dialogue, which proves in his hands a supple medium for the depiction of how philosophical problems emerge from real contexts of thought and action.³⁹ When he gives arguments, he is at pains to ensure their premises are acceptable to his interlocutors, either because they have asserted them themselves, or assented to them, or come to be persuaded of their truth. In this too, one might detect at least a hint of conservatism.

³⁹ This point is argued at length in, e.g., Michael Frede, "Plato's Arguments and the Dialogue Form," ed. James C. Klagge and Nicholas D. Smith, *Methods of Interpreting Plato and His Dialogues* (Oxford Studies in Ancient Philosophy) Supplementary volume (1992): 201–20. Similar views are advanced in Martha C. Nussbaum, *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*, 2d edition (New York: Cambridge University Press, 2001), 87: 'Plato uses the dialogue to motivate a view, to make us feel the force of a problem, to explain the practical roots and implications of a solution'; and in Terence Irwin, "Plato: The Intellectual Background," in *The Cambridge Companion to Plato*, ed. Richard Kraut (Cambridge: Cambridge University Press, 1992), 76: 'Plato chose the dialogue because he thought it stuck most closely to the essential features of Socratic argument. Socrates claims that the systematic, rule-governed form of interrogation that he practices allows him to secure his interlocutors' agreement to moral positions that they would have firmly, often indignantly, rejected before they faced Socrates' questions.'

Now, there is an obvious sense in which his interlocutors are unique and speak for no one but themselves. Yet, they are seldom entirely idiosyncratic. Like human beings in most times and places, a good many of their views, attitudes, more generally their dispositions, are shared with other people. They have, as we should naturally say, *common sense*, that undemanding (but still not ubiquitous) complex of everyday virtues. Their dispositions allow them to get by in much of their everyday lives, including that part of their lives we might call their *ethical* lives, calling this man brave and that one cowardly, this act pious and that other impious, and finding their judgments shared by their fellows. Plato's Socrates subjects these quotidian talents to examination, assessing them by more demanding criteria than those by which they are ordinarily assessed. He does so by asking two sorts of questions about the *terms* in which they exercise these capacities – usually the ordinary-language labels for the recognised virtues, courage, temperance, justice and the like: namely, (1) the question of what they are and (2) the question of whether some or other claims are true of them.

More often than not, what they say in response to these questions is an expression of what may well be termed 'common-sense morality'. No interlocutor of Socrates' is an out-and-out eccentric; even the most provocatively heterodox of his immoralist respondents – Thrasymachus in the *Republic* and Callicles in the *Gorgias* –

have a good many of what the most studiously conformist figure would regard as plausible opinions; this is part of their dangerous appeal. What Socrates produces with his interrogations is a sort of Domesday book of the contemporary ethical common-sense, inevitably incomplete, but with some claim to be broadly representative of some of the more prominent strains in that common sense, from the conformist to the dissenting. So much I take to be relatively uncontroversial; most scholars will accept this as a characterisation of what happens in the dialogues of Plato where Socrates deals extensively with ethical matters. I hope in this essay to offer some further resources for understanding what happens in these dialogues *vis-à-vis* common-sense morality, and explain more thoroughly *how* Plato's Socrates goes about the project of reconstituting this morality along more rational lines.

All subsequent references to Socrates – I perhaps need hardly say – are to the Socrates of the Platonic dialogues: the historical Socrates famously wrote nothing that has reached us; the Socrates in Xenophon and Aristophanes has next to nothing to say on the questions this essay is concerned with. However, I do need to say something about the fact that even Plato's Socrates can be an unpredictable creature, saying things in one dialogue that are at the very least in tension with the things he says in other dialogues. A popular hypothesis that purports to explain

these divergences has acquired the label 'developmentalism'. In its standard scholarly use, it means not simply that Plato's thinking developed over the course of his life (as the thinking of most long-lived thinkers is bound to do) but that at least some of these developments consisted in his changing his mind on fairly important questions, such as the nature and existence of 'Forms', the possibility of acting against one's knowledge of what is best, and the integrity of the soul. The testimony of Plato's near contemporaries is frustratingly ambiguous on this point; consequently, developmentalist claims tend to rest in large part on interpretative theses about the dialogues. Not all developmentalist scholars agree on just what divergences among the dialogues warrant what hypotheses about a change of mind on Plato's part. But an influential line of thought among modern interpreters, with many variations, has it that one set of dialogues is an attempt by Plato to depict the practice of the historical Socrates while another set has Socrates reduced to the status of mouthpiece for distinctively Platonic views on questions about which the historical Socrates said little.

Dialogues such as the *Laches*, *Charmides*, and *Euthyphro* have in common the fact that they are in considerable part concerned with understanding the nature of some virtue, and that they tend to end with Socrates' interlocutors 'at a loss', in a state of *aporia*. These stand in sharp contrast to a dialogue such as the *Republic*, where

Socrates and his interlocutors rise, from the puzzlement into which they fall early in the dialogue, to develop a startlingly detailed picture of moral psychology, politics, metaphysics and epistemology that few interpreters would wish to attribute to the historical Socrates. There are then the detailed excursions into questions of language, logic and philosophical method in dialogues such as the *Parmenides*, *Sophist* and *Statesman*, cosmology in the *Timaeus*, and a different, more elaborately legislative, approach to politics in the *Laws*. A standard view on chronology has emerged, familiar as the default position even to scholars who reject it: there are the early ‘aporetic’ or ‘Socratic’ dialogues, the constructive middle-period dialogues, then the methodologically self-conscious and systematic late dialogues.⁴⁰ The existence of at least that latter category is to some degree corroborated by independent testimonial evidence and modern stylometric techniques.

Scholars who aim to defend a substantive thesis on questions of Platonic biography rarely have space to say much else. I had best simply state my assumptions in what follows. I assume, as most scholars do, that there is a rough pattern to be discerned in the practice of Socrates in the ‘early’ dialogues – a pattern consisting in the

⁴⁰ The literature on this question is vast, in part because every scholar needs to take some position on the chronological question, however conciliatory. Terence Irwin, *Plato's Ethics* (New York: Oxford University Press, 1995), 11–13 has a brief account of what I take to be the standard approach to this question.

recurrence of certain formal features as well as the expression of certain distinctive (sometimes counterintuitive) ethical views. The Socrates of these dialogues rarely theorises his own practice, but there do seem to exist enough in the way of passing remarks and incipient suggestions to comprise the materials for such a theory.

'Later' dialogues (such as the *Republic*, *Theaetetus*, *Sophist* and *Philebus*) supply the basic materials such a theory, materials that go beyond mere formal unity, in particular an epistemology and a psychology. In these dialogues, we find claims about the nature of belief, thought, and knowledge, that offer us a way of understanding what Socrates is doing in the 'early' dialogues, and how and why he does it. (I shall hereby dispense with the inverted commas around the words late and early; readers with unitarian sympathies may read the labels in terms of their now-conventional extensions – 'dialogues conventionally regarded as early/late'.)

I have stated my assumptions in minimal terms compatible with a wide range of chronological views. Certainly they rest on assuming a certain degree of unity and continuity in Plato's intellectual career – otherwise nothing said in one dialogue could permissibly be used to illuminate something said in another without considerable argument.⁴¹ But I assume no more unity than my subsequent arguments need. I do not claim, for instance, that Plato – *a fortiori* the historical

⁴¹ See Catherine Osborne, "Socrates in the Platonic Dialogues," *Philosophical Investigations* 29, no. 1 (2006): 1–21 for an attempt to present the case against trans-dialogue readings.

Socrates – had always possessed the epistemology and psychology of the late dialogues, only that nothing in the latter undermines Socratic practice as depicted in the former. I do, however, think it entirely plausible to suppose that the late dialogues give us grounds for redescribing the scope, nature, and achievements of Socratic practice.

A brief map of what is to follow: in sections 2 – 6, I shall consider the psychological presupposition of Socrates' instruction to his interlocutors to 'say what [they] believe'. I shall find in the dialogues evidence of two possible theories we might attribute to Socrates. On one, Socrates is an *avant la lettre* 'Cartesian' who thinks self-knowledge, in particular with respect to the self-ascription of *beliefs*, can be attained by a process of introspection that reveals pre-existing beliefs in the mind that may then be expressed in dialogue; call this the 'theoretical' picture of self-knowledge, where the interlocutor must ask himself *what it is he believes*. This can be contrasted with what we may call a 'deliberative' picture of self-knowledge, where the interlocutor answers the question of what he believes by asking himself *what he is to believe*. I shall argue for the deliberative interpretation of Socrates' request: what Socrates is asking his interlocutors to do, I claim, is not to find out what they always thought, but to *make up their minds*, and further, to be prepared to speak their minds and take responsibility for what they assert. Not only does this interpretation make

better sense of the relevant passages of text, I claim also that it gives us a better way of understanding the potential therapeutic benefits of Socratic practice, and reconstruct it as a distinctive variety of ‘language game’ that simulates conditions conducive to the acquisition of certain ethical and intellectual virtues. In section 7, I apply elements of this picture of Socratic practice to passages where Socrates is engaged in what we might call social criticism, and offer a reconstruction of the first two books of the *Republic* in their light.

Some readers might find, even in this brief précis, grounds for concern with imminent anachronisms. Can it ever be acceptable to use distinctions from Wittgenstein and 21st-century debates in the philosophy of mind in an essay purporting to attribute philosophical claims to a thinker from Greek antiquity?

There are radical replies available – for instance, to reject what is sometimes called the ‘realist’ assumption in interpretations of texts from the (early) history of philosophy,⁴² but I do not wish to go quite so far. To some degree, any 21st-century

⁴² For a discussion of ‘realism’ in this context, see Nicholas D. Smith, “Editor’s Afterword: Platonic Scholars and Other Wishful Thinkers,” ed. Nicholas D. Smith and James C. Klagge, *Methods of Interpreting Plato and his Dialogues*, *Oxford Studies in Ancient Philosophy*, Supplementary volume (1992): 245. For a pragmatic view on the question, see Bernard Williams, *Shame and Necessity* (Berkeley: University of California Press, 1993), 14: ‘in the difficult case of Plato, it is not that there is anything necessarily wrong in treating texts as the historian of philosophy typically treats them. It is wrong only if it turns out to be unrewarding, or if there is too dominating an idea that we are uncovering the argument that is “really there”. If it is freed from obsession, the activity can be creative and illuminating.’ I have some sympathy for Williams’s position on the history of philosophy – a position developed at greater length in Bernard Williams, “Descartes and the Historiography of Philosophy,”

essay in English about an ancient Greek thinker will be guilty of describing him in terms he would not recognise. Indeed, few scholarly works will manage ‘not to attribute to Plato any proposition which cannot be translated into Greek, the language in which he did his thinking. If it cannot be, he cannot have thought that.’

R. M. Hare, the author of this demanding test, continues:

to be on secure ground, if his own words are unclear or ambiguous, the most we can do is to imagine that we have him with us, put to him questions in Greek, and then speculate as to how he might answer them in Greek. If this method is followed, it will be found that many of the distinctions on which, as modern philosophers, we rightly want to insist, pass him by.⁴³

It may be so, but I will not avail myself of this way out. I believe that the basic distinctions I shall make in this essay – my unfamiliar terminology apart – are ones that Plato would both recognise and endorse.

2. ‘Say what you believe’

There are few general claims we can make about the methods of Plato’s Socrates without fear of counterexample. It is safe only to say that they involve Socrates conversing with one or more interlocutors, and that a significant number of his

in *The Sense of the Past: Essays in the History of Philosophy*, ed. Myles Burnyeat (Princeton: Princeton University Press, 2006), 257–64 – but this essay does not assume it. It is, however, certainly interested in ‘putting the philosophy of the past to use in present terms’ by choosing ‘to neglect or overlook, to some extent, the history that lies between that philosophy and the present day, and to reconsider the philosophy in partial independence from its actual influence’ (258). For a brief, and amusing, rejoinder to Williams, see Jonathan Barnes, “The History of Philosophy,” in *Method and Metaphysics: Essays in Ancient Philosophy I* (Oxford: Oxford University Press, 2011), 17–22.

⁴³ R. M. Hare, *Plato*, Reprint edition (Oxford: Oxford Paperbacks, 1982), 25.

utterances are in the interrogative mood. The modern interpretative scholarship is divided on the question of exactly what Socrates is trying to do, and what his practice presupposes in the way of an epistemology and metaphysics.⁴⁴ Lacking any authoritative and systematic statement of ‘methodology’ in these dialogues, scholars have had to reason inductively, from practice and the odd *obiter dictum*, to theory. But the *obiter dicta* on method in the dialogues are not always consistent with each other or indeed with the practice of Socrates in the dialogues. Further, Plato depicts that practice itself as conspicuously various and marked by Socrates’ notorious irony. Plato himself appears to have undergone several shifts of opinion over the course of a long philosophical career, and it is never a simple matter to distinguish between his attempts to represent faithfully the historical Socrates and his attempts to use the figure of Socrates as a vehicle for his own evolving ideas.

However, assuming an underlying unity of presuppositions behind the evident diversity of Socratic practice has proved a fruitful assumption for modern scholarship. The texts give us strong reason to suppose the existence of at least such a thing as a Socratic *paradigm*, albeit one from which considerable departure may be

⁴⁴ In other words, on the question of whether there is such a thing as ‘the Socratic method’. For a representative sample of the divergent opinions on this question, see the essays in Gary Alan Scott, ed., *Does Socrates Have a Method?: Rethinking the Elenchus in Plato’s Dialogues and Beyond* (Philadelphia: Penn State Press, 2002). I am myself inclined to think that Socrates does have a method, but – as this essay will try to show – that it is not adequately characterised in terms of its merely formal features.

observed in any particular dialogue. The paradigm is familiar enough: Socrates asks a question about some virtue, either asking for it to be defined, or for some other affirmation about it. The interlocutor affirms some proposition p in answer to the original call for definition. Further questions reveals him⁴⁵ also to believe some q and r . It is subsequently shown that $\text{not-}p$ may be derived from q and/or r showing his assertions, and therefore presumably also his beliefs, to be inconsistent. The interlocutor finds himself 'at a loss', in a state of *aporia*. At this point he usually leaves, but on occasion, he stays to continue the conversation, seeking a way out of the *aporia*.

The point can always be pressed with respect to any particular Platonic dialogue whether, and to what degree, Socrates' practice in it is consonant with this basic paradigm. Further questions may be raised about Socrates' aims. To what extent are they concerned with the *truth* of the beliefs expressed in the course of his interrogation, independently of other facts about the believer? To what extent are they merely *ad hominem* and, as such, potentially therapeutic, aimed towards the examination of a life and the reform of the character of the interlocutor whose life it is? If Socrates' aims are indeed therapeutic, why the focus on the interlocutor's beliefs to the exclusion of all the other things that make up a life? These are divisive

⁴⁵ I use male pronouns in the following simply because all the interlocutors of Socrates I shall discuss are male.

questions in the scholarship.⁴⁶ There is, however, one point on which commentators are nearly unanimous in agreement,⁴⁷ that Socrates requires of his interlocutors that (1) they must say – assert or assent to – what they actually believe, and (2) they must not say what they do not believe. The two requirements, positive and negative, can be captured with a biconditional formulation: *an interlocutor S should assert that (or assent to) p iff S believes that p*. The existence of such a requirement is usually thought to be what distinguishes Socrates’ (‘elenctic’) practice from the (‘eristic’) exchanges characteristic of his *bêtes noires*, the sophists and rhetoricians.⁴⁸

On one way of marking the distinction, Socrates needs to enforce the rule if his aims, both philosophical and therapeutic, are to be furthered. Indeed, the *prima facie* evidence for Socrates imposing such a rule is considerable.⁴⁹ There are, however, serious questions to be raised about just what the constraint amounts to. What is a

⁴⁶ A succinct account of the difficulties that follow from virtually any interpretation of Socratic practice may be found in Irwin, *Plato’s Ethics*, 29–30. There is a helpful summary of recent scholarly debates on the question of Socratic method in the editor’s introduction to Scott, ed., *Does Socrates Have a Method*.

⁴⁷ An extensive bibliography of scholars who endorse some or other version of this rule may be found in John Beversluis, *Cross-Examining Socrates: A Defense of the Interlocutors in Plato’s Early Dialogues* (Cambridge: Cambridge University Press, 2000), 37 n. 1 and 38 n.2. A more recent bibliography is Hugh Benson, “Socratic Method” in Donald R. Morrison (ed.), *The Cambridge Companion to Socrates* (Cambridge: Cambridge University Press, 2010), 186 n. 24. The most influential ascription of this rule to Plato’s Socrates appears in Gregory Vlastos, *Socrates: Ironist and Moral Philosopher* (Cornell University Press, 1991), 113, where it is called the “say what you believe” rule’.

⁴⁸ See, e.g., Irwin, *Plato’s Ethics*, 21; Alexander Nehamas, “Eristic, Antilogic, Sophistic, Dialectic: Plato’s Demarcation of Philosophy from Sophistry,” *History of Philosophy Quarterly* 7, no. 1 (January 1, 1990): 10.

⁴⁹ E.g. See *Laches* 193c6–8; *Euthyphro* 9d7–8; *Crito* 49c11–d1; *Protagoras* 331c4–d1; *Republic* I.349a4–8; *Gorgias* 438a1–b1, 495a5–9, 499b4–c6, 500b5–c1, *Theaetetus* 154c7–e6 I shall discuss some of these passages at greater length below.

Socratic interlocutor being asked to do when he is asked to express his actual beliefs and disallowed from asserting anything contrary to them?

The scholarship has tended to see the positive and negative statements as the same 'rule' (or constraint, or requirement, or *desideratum*), but Socrates uses different formulations in different places. It is worth making it pedantically explicit that there are in fact two rules here, not one, if only because the latter, negative, statement of the rule does not entail the former, in that it would be consistent with the latter but not the former for the interlocutor to remain silent. Socrates might well think that the two rules go together, but they are certainly logically distinct. However, no interlocutor ever avails himself of the logically permissible option of silence, so the biconditional above seems to capture what we need of Socrates' requirement.

A further point about my own terminology: the two constraints are often thought to constitute a rule of 'sincerity', but I shall avoid this term in what follows for three reasons. Firstly, the term is a commentarial construct that does not correspond directly to any word in Plato's Greek. Secondly, the term imports a range of ethical connotations (of good faith, intent, and so forth) that we should be better off without. While I shall argue that ethical considerations will go into explaining

Socrates' reasons for imposing such a rule when he does, this should be a matter for argument, not something to be smuggled in by terminological fiat. Thirdly, linguistic intuitions about the English word 'sincere' are, as a matter of fact, conflicted, as is evident in modern efforts to give an account of sincerity. What does sincerity involve? Simply saying what one believes? Saying what one knows one believes? Saying what one believes oneself to believe? Things are even more puzzling when one considers linguistic intuitions about 'insincere'. Is one being insincere if one misstates one's beliefs out of carelessness and ignorance? If one states one's true belief, say in an undetected Freudian slip, while intending to deceive: is one being sincere or insincere? These are potentially rewarding questions and amenable to interesting 'ordinary language' treatments, but they seem to me strictly orthogonal to any discussion of Socrates.

The demand to say only what one believes is, it is widely held, essential to Socratic practice.⁵⁰ As Gregory Vlastos has put it, that practice has 'a double objective: to discover how every human being ought to live and to test that single human being

⁵⁰ This is not an uncontroversial point. The passages usually cited as counterexamples to the claim that Socrates *always* imposes such a rule are his exchanges with sophists/rhetoricians where he carries on his procedures while fully aware his interlocutors' assent is ironic or convenient or bored, in any case no longer sincere: e.g. *Protagoras* 332b ff., *Gorgias* 499b4–c6, 501c7–8, 505c5 ff., *Republic* I 349a9 ff. See Thomas C. Brickhouse and Nicholas D. Smith, *Plato's Socrates* (Oxford: Oxford University Press, 1994), 14 n. 21 for a brief defence of this position, and T. H. Irwin, "Say What You Believe," *Apeiron* 26, no. 3/4 (1993): 1–16 for discussion of the interpretative difficulties posed by these passages.

who is doing the answering – to find out if he is living as one ought to live’.⁵¹ As a reviewer of Vlastos puts it, ‘These two objectives can only be pursued together if the interlocutor produces sincere answers and thus exposes his inner world to scrutiny’.⁵² It is this metaphor of the ‘inner world’, strongly suggested by Vlastos’s theory of the Socratic *elenchus* and tacitly accepted in much subsequent scholarship, that I hope to challenge in the following sections of this essay. The metaphor conscripts Socrates into a (roughly) Cartesian model of self-knowledge, derived from some aspects of Plato’s depiction of Socratic practice, that is at odds with a strikingly anti-Cartesian picture also to be found in the dialogues. It is my claim that it is the anti-Cartesian picture that is the more basic and philosophically appealing.

3. Why say what one believes?

‘Say what you believe’, says Socrates; ‘do not say what you don’t believe’. But why *need* he say that? Is that not one of the constitutive norms of conversation, well captured in the Gricean ‘maxim of quality’? What reason might he have for, and what good could be done by, making it explicit? A hypothesis suggests itself: he must make it explicit because he suspects his interlocutors are at risk of violating it.

Whence such a risk? In the *Crito* (49c–e), Socrates tells Crito:

⁵¹ Gregory Vlastos, *Socratic Studies* (Cambridge: Cambridge University Press, 1994), 10.

⁵² Martha C. Nussbaum, “Socratic Studies,” *The Journal of Philosophy* 94, no. 1 (January 1, 1997): 28.

One should never do wrong in return, not do any man harm, no matter what he may have done to you. And Crito, *see that you do not agree to this, contrary to your belief* (*mē para doxan homolegēs*). For I know that only a few people hold this view or will hold it, and there is no common ground (*koinē boulē*; alternatively, ‘common council’) between those who hold this view and those who do not, but they inevitably despise each other’s views. So then consider very carefully whether we have this view in common, and whether you agree (*koinōneis kai sundokei*), and let this be the basis of our deliberation, that neither to do wrong nor to return a wrong is ever correct, nor is doing harm in return for harm done. Or do you disagree and do not share this view as a basis for discussion? I have held it for a long time and still hold it now, but if you think otherwise, tell me now. If, however, you stick (*emmeneis*) to our former opinion, then listen to the next point.

Crito: I stick to it (*emmenō*) and agree (*sundokei moi*) with you. So say on. (*Crito* 49c–e, emphasis added)⁵³

Socrates’ reasons for making the rule explicit – here stated in its negative form – emerge naturally from the context. The belief in question is unorthodox; not many people hold it. All the more reason, then, to confirm that Crito is not just nodding along.⁵⁴ Further, the situation is fraught: Crito is talking to a Socrates facing imminent execution. What could be more natural than for him to concede the point out of politeness or consideration for Socrates? Socrates needs to make sure Crito is not simply indulging him his eccentric opinion.

⁵³ This, and all subsequent quotations from Plato unless otherwise specified, are from Plato, *Complete Works*, ed. John M. Cooper and D. S. Hutchinson (Indianapolis: Hackett Publishing, 1997). I shall indicate my own modifications and emphases where necessary.

⁵⁴ See Irwin, *Plato’s Ethics*, 32 for a brief discussion of the contrast between this passage and the many where the Socratic interlocutor agrees with a widely shared view. A similar view is defended in Brickhouse and Smith, *Plato’s Socrates*, 15 n. 26.

Another sort of exchange occurs in the *Protagoras*, where Socrates asks Protagoras whether he agrees that justice and piety are the same sort of thing. Protagoras

replies:

It's not so absolutely clear a case to me, Socrates, as to make me grant that justice is pious, and piety is just. It seems a distinction is in order here. But what's the difference? If you want, we'll let justice be pious and piety just.

Socrates won't have it:

Don't do that do me! It's not this 'if you want' or 'if you agree' business I want to test. It's you and me I want to put on the line, and I think the argument will be tested best if we take the 'if' out. (*Protagoras* 331c)

Here, Protagoras, one of Socrates' more intelligent and self-aware interlocutors, is quite understandably inclined to violate the rule because he cannot get himself to assent to Socrates' thesis in its unqualified form; as he puts it, 'a distinction is in order here'. But if Socrates wants assent to a simple thesis, he will provide it while retaining private reservations. However, Socrates makes it clear that he wants more than just hypothetical assent. The motivation to explicate the rule is Protagoras's suggestion – expressed in his 'if you want' – that he is violating it, assenting to something he does not (quite) believe.

A different motivation for explicating the rule appears in the exchange with Thrasymachus in the *Republic*: 'we mustn't shrink from pursuing the argument and looking into this, just as long as I take you to be saying what you really think. And I

believe you aren't joking now, Thrasymachus, but are saying what you believe to be the truth' (*Republic* 349a). Here, Socrates is making it clear that an ironic or mocking response will not do; there is not much point trying to refute a position advanced in jest.

Yet another motivation appears in the *Theaetetus*, where Socrates is describing to the titular interlocutor the structure of a familiar sophistical manoeuvre, but in a manner that reveals an uncanny, if superficial, resemblance to his own methods:

... Here are six dice. Put four beside them, and they are more, we say, than the four, that is, half as many again; but put twelve beside them, and we say they are less, that is, half the number. And there is no getting out of that (*oude anekton allōs legein*; or, 'there is nothing else admissible to say') – or do you think there is?

No, I don't.

Well now, supposing Protagoras or anyone else were to ask you this question: 'Is it possible, Theaetetus, for any thing to become bigger or more in number in any way other than by being increased?' What is your answer to that?

Well, Socrates, if I answer what seems true in relation to the present question (*to dokoun pros tēn nun erōtēsīn*), I shall say 'No, it is not possible'; but if I consider it in relation to the question that went before, then in order to avoid contradicting myself (*phulattōn mē enantia eipō*), I say 'Yes, it is.'

That's a good answer, my friend, by Jove it is; you are inspired. But, I think, if you answer 'Yes', it will be like that episode in Euripides – the tongue will be safe from refutation but the mind will not⁵⁵ (*hē men gar glōtta anelenktos hēmin estai, hē de phrēn ouk anelenktos*).

That's true.

⁵⁵ The allusion is to *Hippolytus* 612 where Euripides has Hippolytus say 'My tongue has sworn the oath, my mind has not' (*hē glōss' omōmoch', hē phrēn anōmotos*). See the discussion of the context and significance of the utterance in Jon D. Mikalson, *Honor Thy Gods: Popular Religion in Greek Tragedy* (UNC Press Books, 1992), 84–6.

Now if you and I were professional savants, who had already analyzed all the contents of our minds (*panta ta tōn phrenōn exētakotes*), we should not spend our superfluous time trying each other out; we should start a regular Sophists' set-to, with a great clashing of argument on argument. But, as it is, we are only plain men; and so our first aim will be to look at our thought themselves in relation to themselves, and see what they are (*theasasthai auta pros hauta ti pot' estin ha dianoometha*) – whether, in our opinion, they agree with one another (*allēlois sumphōnei*) or are entirely at variance.

That would certainly be my aim, anyway.

And mine. That being so, as we are not in any way pressed for time, don't you think the thing to do is to reconsider this matter quietly and patiently, in all seriousness 'analyzing' ourselves (*hēmas autous exetazontes*), and asking what are these apparitions (*ta phasmata*) within us? [...]

Socrates proceeds to conduct such an analysis, coming up with three such 'apparitions', then says:

Now it seems to me that these three statements that we have admitted (*hōmologēmata*; or, 'assumptions') are fighting (*machetai*; or, 'wrangling', 'contending [with]') one another in our souls [...] (*Theaetetus* 154c–155b)

The motivation for Theaetetus to violate the rule faced with Protagorean tactics is an awareness that to say what one believes would be to open oneself up to the charge of inconsistency. And certainly, as Socrates allows, one might avoid the charge by violating the rule. But that would be a Pyrrhic victory, consistency in assertion without consistency in thought.⁵⁶

⁵⁶ Examples may be multiplied, but one more, from the *Laches* (193c), will suffice. 'Why, what else would anyone say, Socrates?' says Laches. 'Nothing [else], if that is what he thought', replies Socrates. As Irwin, *Plato's Ethics*, 20 notes, 'if [Socrates' interlocutors] altered their claims simply to maintain or to restore consistency, they would be violating a principle of the elenchos ... [W]hen Laches agrees that he does believe it, he gives the answer even though it causes trouble for his initial claim'.

The passages above suggest various possible explanations for Socrates mentioning his rule explicitly, explanations based on the tendency of certain of his interlocutors to misreport the content of their beliefs either out of some sense of *politesse*, or a misunderstanding of the sort of conversation Socrates seeks to have, or deliberate irony, or a desire for consistency in their assertions. At least in some of these cases, the rule, though it is stated as a requirement, seems to function rather as a permission of sorts: ‘don’t worry, you don’t have to please me, or humour me, or entertain me, or impress me with your consistency; just say what you think’.

There is, however, one thing that unites all these explanations: Socrates seems to assume that (1) his interlocutor has some belief on the matter, and that (2) it is *up to the interlocutor* to state or not to state that belief. (1) seems a safe assumption: The possibility that an interlocutor is affecting to have a belief when in fact he has no belief on the matter is not mentioned explicitly in any Platonic dialogue; the subjects about which Socrates interrogates his interlocutors tend to be subjects on which they do have beliefs. In any case, this eventuality is covered by the rule: for S to assert or assent to *p* when S in fact has no belief whether or not *p* would be for S to assert or assent to *p* when it is not the case that S believes that *p*. The more interesting case is when an interlocutor does have a belief on the subject Socrates is asking him about, as Socrates is assuming, but either (1) does not know what it is or

(2) does not know how best to state it. If they violate the rule in these circumstances, it is not out of a desire to dissemble.

(1) and (2) are related but distinct cases. The first of them is a matter of self-knowledge; we have a good many beliefs but not all of them are present to, or even easily available to, the consciousness. The second is a matter of articulacy: we have a good many complex (subtle, nuanced) beliefs to whose complexity we are not always able to do justice. Some beliefs are dispositional, and not all our dispositional beliefs are present to the immediate consciousness; nor do they come ready-articulated. While some dispositional beliefs – take Ryle’s classic example⁵⁷ of the belief that ‘the ice is dangerously thin’ – will be manifested in a wide variety of ways (a proneness to avoid the ice, or to step gingerly when on it, or to keep one’s children off it), they nevertheless admit of articulation in a simple English sentence. Other dispositional beliefs might not allow for such succinct articulation. When the dispositional *explananda* are sufficiently complex, a simple sentence will prove inadequate. Rather one will have to look to something more complex, likely a whole network of reinforcing propositions. To perform this feat of articulation would count as something of an intellectual *achievement*, and not everyone is capable of

⁵⁷ Gilbert Ryle, *The Concept of Mind* (Chicago: University of Chicago Press, 1949), 135.

it.⁵⁸ In other words, if beliefs are understood in terms of their function, then it is entirely intelligible that even the cleverest might sometimes fail to produce an articulation adequate to the *explananda*; this inarticulacy will, *a fortiori*, afflict Socrates' interlocutors. Indeed, this ('functionalist') way of thinking about belief – where the belief expressed by a single, often simple, sentence is to be identified with a state of a person that performs a complex functional role, involving patterns of inference, ontological commitments, and dispositions of behaviour, assertion and assent – is the closest thing to an orthodoxy in contemporary philosophy of mind.

4. Forming beliefs

A revealing passage early in the *Charmides*, seldom discussed in this connexion, is worth reading closely for what it presupposes about belief. The young and beautiful Charmides has been suffering headaches in the morning (it is at least hinted that these are hangovers), and is invited by his uncle Critias to converse with Socrates on the understanding that Socrates might be able to cure him of them. Socrates starts by asking him if he deserves the reputation Critias says he has, of 'being the most temperate young man of the day' (*Charmides* 157d): 'So tell me yourself: do you

⁵⁸ One aspect of the difficulty of articulation is the demand for generality – some one single thing to unite a disparate body of *explananda*. As Nussbaum, *Fragility of Goodness*, 10 puts it: 'as Socrates showed ..., most people when asked to generalize, make claims that are false to the complexity and the content of their actual beliefs. They need to learn what they really think.'

agree with your friend and assert that you already partake sufficiently of temperance, or would you say that you are lacking in it?' Charmides blushes bashfully and points out that Socrates' question has placed him in an awkward position: to say yes would be arrogant and to say no would be disrespectful to those whose opinions have ascribed such a reputation to him in the first place. Socrates sees his point:

'... What you say appears to me to be reasonable, Charmides. And I think ... we ought to investigate together the question whether you do or do not possess the thing I am inquiring about, so that you will not be forced to say anything against your will (*mēte su anankazē legein ha mē boulei*) and I, on the other hand, shall not turn to doctoring in an irresponsible way. If this is agreeable to you, I would like to investigate the question with you, but if not, we can give it up.'

'Oh, I should like it above all things,' he [Charmides] said, 'so go ahead and investigate the matter in whatever way you think best.'

'Well then, ... in these circumstances, I think the following method would be best. Now it is clear that if temperance is present in you, you have some opinion (*doxazein*) about it. Because it is necessary, I suppose, that if it really resides in you, it provides a sense (*aisthēsin*) of its presence, by means of which you would form an opinion (*doxa*) of what it is and of what sort it is. Or don't you think so?'⁵⁹

'Yes,' he said, 'I do think so.'

'Well, then, since you know (*epistasai*) how to speak Greek, ... I suppose you could express this impression of yours in just the way it strikes you?' (*kan eipois dē pou auto hoti soi phainetai*)

⁵⁹ I have modified the Hackett translation by Rosamund Kent Sprague slightly. Sprague has 'by means of which you would form an opinion not only that you have it but of what sort it is' for 'ex hēs doxa antis soi peri autēs eiē hoti estin kai hopoion ti'. This seems to me linguistically tenable, but problematic once the rest of the passage is taken into account. If the line means what Sprague takes it to mean, Socrates is claiming that to possess temperance implies having an opinion (*doxa*) that one possesses it. But the next lines suggest that Socrates wishes precisely to determine whether he has it or not, and to do so by having him express an opinion about *what it is*. If his opinion can withstand Socratic cross-examination, then there might be grounds for saying that he does possess it. There would not be the same motivation for the demand to define temperance if Charmides already has an opinion about its presence in him (or otherwise).

‘Perhaps (*isōs*),’ he said.

‘Well, to help us decide (*topasōmen*; alternatively, ‘to divine’) whether it resides in you or not, say what, in your opinion (*kata tēn sēn doxan*), temperance is ...’ (158d–159a)

As in other places, Socrates assumes without argument that Charmides does have *some* belief about what temperance is, a belief that Charmides can draw on in response to his request for a definition. In general, when Socrates asks his interlocutors for a definition of some virtue, his motivation for his doing so typically arises from something in the preceding conversational exchange.

Sometimes it is an explicit claim of knowledge – as with Euthyphro who claims to ‘have accurate knowledge of all such things’ (*Euthyphro* 5a1), that is to say of piety and impiety. A knowledge claim is (*a fortiori*) a belief claim. Sometimes it is the fact that one has used an undefined virtue term in conversation and thus taken some conception of that virtue for granted: *Meno* presupposes some antecedent conception of virtue, however minimal or untheorised, simply in asking Socrates whether it can be taught or not (*Meno* 70a1); the remarks of old Cephalus in *Republic* I (330d–331b) associate justice with truthfulness and solvency. With Charmides, however, it is not just that he has *mentioned* temperance; he has also claimed to possess it, or at any rate failed to disavow a claim to that effect made on his behalf.

Socrates’ remarks suggest more. He seems also to think the expression of belief will take some effort on Charmides’ part – some combination of reflection and an

attempt at linguistic articulation. It is worth pausing a moment on Socrates' phrase, 'since you know how to speak Greek'. He might well have said, as he presumably means to say, 'since you can speak', or more colloquially, 'since you've got a tongue in your head'. But we should entertain the thought that Plato might have meant the specific reference to knowing Greek to call to mind the other place in the Platonic corpus where Socrates brings up an interlocutor's knowledge of Greek. When Meno asks him to demonstrate the intriguing 'theory of recollection' he has just set out, Socrates replies:

It is not easy, but I am nevertheless willing to do my best for your sake. Call one of these many attendants of yours, whichever you like, that I may prove it to you in his case.

Certainly. You there come forward.

Is he a Greek? Does he speak Greek?

Very much so. He was born in my household. (*Meno* 82ab)

Socrates' question to Meno is not really about ethnicity but language, and Meno rightly understands it as such.⁶⁰ It would have been a real question about any given slave whether he spoke Greek or not, given the frequent involvement of Greek militaries in campaigns among foreign ('barbarian') peoples in the 5th century.

Socrates wants to know whether he can carry on a conversation with the slave, one

⁶⁰ For a discussion of the relationship between Greek ethnic self-conception and the Greek language, see Edith Hall, *Inventing the Barbarian: Greek Self-Definition Through Tragedy* (Oxford: Oxford University Press, 1989), 5: 'No other ancient people privileged language to such an extent in defining its own ethnicity', and *passim*.

that will involve having him assert and assent to a wide range of (as it happens, geometrical) claims. If the slave turns out to know no Greek, or does not know it well enough to participate in such a conversation, it would not of course follow that he has no beliefs. However, what beliefs he has can only be made public, shared, discussed by linguistic means shared by the parties to the discussion. Given the subject matter, the language in question must be able to make a fair number of subtle distinctions if it is to be adequate to the conversation they are about to have.⁶¹

Socrates is evidently drawing a connexion between belief and language, though there is not enough textual material to specify exactly how he conceived of that connection. There is, however, some material on how he conceived of the relation between thought, belief and assertion, in the late dialogues. A central text is in the *Theaetetus*, where Socrates defines thinking (*dianoesthai*) as:

A talk (*logon*) which the soul has with itself about the objects under its consideration. Of course, I'm only telling you my idea in all ignorance; but this is the kind of picture I have of it. It seems to me that the soul when it thinks is simply carrying on a discussion (*dialegesthai*) in which it asks itself questions and answers them itself, affirms and denies (*phaskousa kai ou phaskousa*). And when it arrives at something definite (*horisasa*), either by a gradual process or a sudden leap, when it affirms one thing consistently and without divided counsel, we call this its judgment (*doxan*). So, in my view, to judge (*to doxazein*; alternatively, 'to believe'

⁶¹ It is of course possible that Socrates shared what Hall, *Barbarian*, 11 describes as 'the Greeks' sense of their own superiority' and thought no language but Greek adequate to this sort of exchange, but his question in the *Meno* is too thin a basis for such an inference. (His remarks on the Cretan and Spartan origins of philosophy in the *Protagoras* 342ab do not offer much more in the way of evidence.)

or ‘to opine’)⁶² is to make a statement, and a judgment is a statement which is not addressed to another person or spoken aloud, but silently addressed to oneself (189e6–190a6; similarly, *Sophist* 263e3–4a2)

The connexion between this silent dialogue and public assertion is drawn in the *Philebus*, in Socrates’ exchange with Protarchus about the relation between pleasure and true or false judgment, an exchange that is well worth quoting in full:

Wouldn’t you say that it often happens that someone who cannot get a clear view because he is looking from a distance wants to *make up his mind* about (*krinein*; literally, ‘distinguish’) what he sees?⁶³

I would say so.

And might he then not again raise another question for himself? ... “What could that be that appears to stand near that rock under a tree?” – Do you find it plausible that someone might say these words to himself when he sets his eyes on such appearances?

Certainly.

⁶² The (strong) reasons for preferring the translation ‘to judge’ are well summarised in Myles Burnyeat, “Introduction”, *The Theaetetus of Plato* (Indianapolis: Hackett Publishing, 1990), 69: ‘First, *doxa* is initially described as an experience (187d) or episode (187e) occurring at a particular time and place. [It is later] defined as a statement made silently to oneself (190a). All this is better suited to “judgement” than to “belief”, which stands for a continuing state of mind or, on some philosophical views, a disposition. Second, although “opining” can, unlike “believing”, refer to an act or episode, “opinion” imports a *contrast* with “knowledge” and would make Theaetetus’ second definition [of knowledge as “true *doxa*”] seem absurd rather than merely false.’ On this question, see the broadly similar views in David Sedley, *The Midwife of Platonism: Text and Subtext in Plato’s Theaetetus* (Oxford: Oxford University Press, 2004), 118–9; David Bostock, *Plato’s Theaetetus* (Oxford: Oxford University Press, 1988), 156–7.

⁶³ There is certainly an interpretative choice involved in rendering ‘*krinein*’ as ‘to make up [one’s] mind’, but the choice is intelligible in the light of the *Theaetetus* passage quoted above. We might also compare an earlier *Theaetetus* passage where Socrates characterises the formation of a *doxa* – belief/judgment – as a case where ‘the soul itself *attempts to reach a decision* (*krinein peiratai hēmin*) for us by rising to compare them with one another’ (186b). It is true that these translational choices support the view I am shortly to advance about Plato’s view of belief formation being (at least analogous) to practical decision, but my view does not rest on my reading into Plato’s Greek connotations only present in the English; rather, as I shall argue, the appeal of such ‘deliberative’ phrasing is a consequence of an idea in Plato’s text to which translators have instinctively, and correctly, responded.

And might he not afterwards, as an answer to his own question, say to himself, “It is a man,” and in so speaking, would get it right?

No doubt.

But he might also be mistaken and say that what he sees is a statue, the work of some herdsmen?

Very likely.

But if he were in company, he might actually say out loud to his companion what he had told himself, and so what we earlier called judgment (*doxan*) would turn into an assertion (*logos*)?

To be sure.

Whereas if he is alone, he entertains this thought by himself, and sometimes he may even resume his way for quite a long time with the thought in his mind?

No doubt.

But look, do you share my view on this?

What view?

That our souls in such a situation is comparable to a book?

How so?

If memory and perceptions (*aisthēsesi*) concur with other impressions (*ta pathēmata*) at a particular occasion, then they seem to me to inscribe (*graphein*; literally, ‘scratch’, ‘etch’) words (*logous*) in our soul, as it were. And if what is written is true, then we form a true judgment (*doxa ... alēthēs*) and a true account (*logoi ... alētheis*) of the matter. But if what our scribe writes is false, then the result will be the opposite of the truth. (*Philebus* 38c–39a)

Reading these two passages together,⁶⁴ we have the basic materials for a Platonic theory of belief formation, at any rate one that is adequate to an analysis of what it

⁶⁴ At this point, it might be worth returning to an objection briefly considered earlier, that one ought not to read any of the sophisticated epistemology of the ‘late’ dialogues into Plato’s depictions of Socrates’ relatively unselfconscious and ethically-oriented practice in the ‘early’ ones. This objection has some force when it combines a strong claim about the extent of Plato’s ‘development’ between the two sets of dialogues (the *Theaetetus* is generally thought to be a dialogue of Plato’s late-middle or

is that Socratic questioning elicits from a typical interlocutor in the ‘Socratic’ dialogues. To form a belief, most conspicuously so in cases where forming a belief requires effort and is more than just a matter of spontaneous assertion (I shall have more to say on the latter case in due course), involves a series of steps: reflection, questioning, ambivalence, perhaps vacillation, then an attempt to resolve one’s ambivalence by settling, non-arbitrarily, for one side of the ambivalence, and then articulating what one has settled for. Articulation is the last step of a process that begins in reflection, and may – though it need not – be followed by and vocalised in an assertion. The articulation, in this picture, must be *in words*, presumably the words of some natural language.⁶⁵ An articulation of this sort produces a sort of *calcified* belief, etched (to switch the metaphor) into the soul beyond easy erasure.

early-late period), and the further claim that the early dialogues strive to present a more-or-less accurate picture of the historical Socrates. But even if these claims are conceded for argument’s sake, it does not strike me as a good reason not to read these passages together. Even if Socrates, either the historical Socrates or the Socrates of the early dialogues to the extent that he is a Platonic creation, had no explicit epistemology underlying his practice, there is enough consistency in that practice to justify the ascription to him of an implicit epistemology, one that the later dialogues explicate, develop and (to some degree) defend. Certainly we should not assume that the epistemology in the later dialogues will vindicate *everything* that Socrates does in the early dialogues, but the burden of argument is on anyone who denies the legitimacy of treating these dialogues as continuous to show why some particular change of view on Plato’s part means that we can no longer treat the explicit epistemology as providing a basis for Socrates’ naïve practice. A more extensive account would be desirable, but beyond the scope of this essay, if only because it would require a fuller account of the relation between belief and the Forms, entities that are not prominent in either the early dialogues or the *Theaetetus*, but are central to the *Republic* and its epistemology in Books V–VII. A discussion of the problems in relating these dialogues appears in Christopher Bobonich, *Plato’s Utopia Recast: His Later Ethics and Politics* (Oxford: Clarendon Press, 2004), 383.

⁶⁵ I cannot see any grounds for ascribing to Plato a view about ‘mentalese’, such as that defended in Jerry A. Fodor, *The Language of Thought* (Cambridge, Mass.: Harvard University Press, 1975). On a related point, Bostock, *Plato’s Theaetetus*, 124 argues that the 189e–190a passage about belief formation ‘shows that [Plato] took one of the prerequisites [of reaching judgement] to be mastery of

This is a brisk reconstruction of a brisk statement of a position for which Socrates provides little argument. Indeed, he protests his ignorance even more insistently than usual, making no further claim for this picture than that this is how things appear to him ('Of course, I'm only telling you my idea in all ignorance; but this is the kind of picture I have (*moi indalletai*) of it'). Keeping this picture in mind, we might return to the *Charmides* passage from which we digressed and find new illumination. When Charmides admits that his first definition of temperance as a sort of quietness has been refuted, Socrates says:

'Then start over again, Charmides ... and look into yourself (*eis seauton emblepsas*), with greater concentration, and when you have decided what effect the presence of temperance has upon you and what sort of thing it must be to have this effect, then put all this together and tell me clearly and bravely, what does it appear to you to be?' (*Charmides* 160de)

The injunction to 'look into [one]self' (*emblepsas*, 160d6) is naturally interpreted as a call to introspect.⁶⁶ Indeed, the Latinate 'introspect' looks like a literal translation of the Greek word. However, I have so far avoided the term, preferring 'reflection', for

a language. Perhaps ... the long and troublesome education that he refers to [at 186c] is the process of learning to talk. After all, this does take time, and only humans can do it'; this is a plausible suggestion, as long as *mastery* is not thought to require more than a working knowledge of a language sufficient to mark the distinctions between which the soul oscillates before reaching a resolution.

⁶⁶ Recall Heraclitus' mysterious fr. 101: '*edizēsamēn emēōton*'; see also the brief discussion in W. K. C. Guthrie, *A History of Greek Philosophy: Volume 1, The Earlier Presocratics and the Pythagoreans* (Cambridge University Press, 1962), 419: 'by the two words of fr. 101 Heraclitus meant, I suggest, first, "I turned my thoughts within and sought to discover my real self"; secondly, "I asked questions of myself"; thirdly, "I treated the answers like Delphic responses hinting, in a riddling way, at the single truth behind them, and tried to discover the real meaning of my selfhood; for I knew that if I understood my self I would have grasped the *logos* which is the real constitution of everything else as well'. The contrast between this sort of picture and the one I am attributing to Plato's Socrates will be evident.

two reasons, linguistic and philosophical. Socrates' verb does not have any special first-personal sense, and he uses cognate expressions elsewhere (155c8: *eneblepsen te moi*; 162b11: *eis ton Kritian apeblepen*) to refer to something quite unlike introspection.⁶⁷ The more interesting reason to avoid introspective language is this. If Charmides is really being called on to introspect, then the material for introspection in this case is phenomenological: Charmides is to introspect on 'what effect the presence of temperance has upon [him]'. But this raises an obvious difficulty: as the dialogue hints subtly, Charmides is not temperate (hence the hangovers alluded to in the conversational exchange that precedes the philosophical discussion, and hence perhaps his inability to define temperance); Socrates' references to his temperance are evidently laced with his characteristic irony. But if Charmides lacks temperance, there is no distinctive effect it produces on him, and thus no special phenomenology for his introspection to fix on. Nevertheless, he does manage to articulate something. What is it?

The most natural conjecture is that he is articulating – poorly – some *conception* of the virtue of temperance he has acquired not from possessing the virtue itself, but rather from living in a milieu and amid a discourse rife with references to it.⁶⁸ Note

⁶⁷ This point is well made in Raphael Woolf, "Socratic Authority," *Archiv Für Geschichte Der Philosophie* 90, no. 1 (2008): 22.

⁶⁸ See Irwin, *Plato's Ethics*, 21 for a brief gloss of this passage.

the description of Charmides' first attempt at definition, attending in particular to his qualifiers and the hints at his reluctance and effort:

At first he shied away (*ōknei*) and was rather unwilling (*ou panu ēthelen*) to answer. Finally (*epeita*), however, he said that in his opinion (*hoi dokoi*) temperance was doing everything in an orderly and quiet way – things like walking in the street, and talking, and doing everything in a similar fashion. 'So I think (*moi dokei*),' he said, 'taking it all together (*sullēbdēn*), that what you ask about is a sort of quietness.' (*Charmides* 159b)

The second time around, he appears to take a better attitude to the whole business:

He paused (*epischōn*) and, looking into himself (*diaskepsamenos*) very manfully, said, 'Well, temperance seems to me (*dokei ... moi*) to make people ashamed and bashful, and so [...]'⁶⁹ modesty must be what temperance really is.' (*Charmides* 160e)

From the shrinking reluctance of his first attempt at reflection, we find an altogether more engaged ('manful') interlocutor. Compare this with the almost exactly parallel advance that Theaetetus makes. Faced with the same demand for a linguistic articulation of what knowledge is ('I can't get a proper grasp of what on earth knowledge really is. Could we manage to put it into words?'), asks Socrates at 146a), Theaetetus too is disinclined to put forward any authoritative claims. In many respects the opposite of Charmides – nowhere as handsome but more obviously inclined to intellectual pursuits (143e–144b) – his first efforts too are stumbling:

... I assure you Socrates, I have often tried to think this out, when I have heard reports of the questions you ask. But I can never persuade myself that anything I say will really do; and I never hear anyone else state the matter in the way that you require. And yet, again, you know, I can't even stop worrying about it.

⁶⁹ The ellipsis here leaves out an 'I think' in Sprague's translation that is not in the Greek, though it is of course implicit in the 'seems to me' earlier in the sentence, but this is captured by Sprague's 'must be'.

Yes; those are the pains of labor, dear Theaetetus. It is because you are not barren but pregnant.

I don't know about that, Socrates. I'm only telling you what's happened to me. (*Theaetetus* 148e)

Socrates responds with that memorable development of his passing reference to the pains of labour, his conception of himself as engaged in a sort of intellectual analogue to midwifery:

The difference is that I attend men and not women, and that I watch over the labor of their souls, not of their bodies. And the most important thing about my art is the ability to apply all possible tests to the offspring, to determine whether the young mind is being delivered of a phantom (*eidōlon*), that is, an error, or a fertile truth. (*Theaetetus* 150bc)⁷⁰

The elaboration of the analogy ends with this gendered peroration:

So begin again, Theaetetus, and try to say what knowledge is. And don't on any account tell me that you can't. For if God is willing, and you play the man (*andrizeē*), you can. (*Theaetetus* 151d)

Having heard Socrates' hortatory remarks, Theaetetus answers with a better grace – nothing in Plato's characterisation suggests that his reply contains any sarcasm – but note the persistence of qualifying locutions:

Well, Socrates after such encouragement from you, it would hardly be decent (or, 'it would be indecent or shameful', *aischron*) for anyone not to try his hardest (*prothumeisthai*) to say what he has in him. Very well then. It seems to me (*dokei ... moi*) that a man who knows something perceives what he knows, and the way it appears at present, at any rate (*hōs ge nuni phainetai*), is that knowledge is simply perception.

That's a good *frank* (*gennaiōs*; alternatively, 'well-born', 'noble') answer, my son. That's the way to speak one's mind (*apophainomenon legein*). (*Theaetetus* 151de; emphases added)

When this definition is refuted (at 186e), Socrates invites him to have a second try:

⁷⁰ See the discussion of this passage in M. F. Burnyeat, "Socratic Midwifery, Platonic Inspiration," *Bulletin of the Institute of Classical Studies* 24, no. 1 (December 1, 1977): 7–16.

... Now look back to the beginning. Wipe out all that we have said hitherto, and see if you can see any better (*ei ti mallon kathorais*) from where you have now progressed to. Tell me again, what is knowledge?

Well, Socrates, one can't say that it is judgment in general, because there is also false judgment – but true judgment may well be knowledge. So let that be my answer (*kai moi touto apokekrithō*). If the same thing happens again, and we find, as we go on, that it turns out not to be so, we'll try something else.

And even so, Theaetetus, you have answered me in the way one ought – with a good will (*prothumōs*), and not reluctantly (*ōkneis*), as you did at first. (*Theaetetus* 187a–c)

The initial reluctance, the sense of inadequacy, is gone. There is no sign of irony in Socrates' words, and indeed, Theaetetus is the finest of all his young interlocutors by far: not for him the warrantless claim to knowledge that make Euthyphro and Laches such difficult interlocutors. However, he is set apart even from Charmides – who shares his disinclination to claim knowledge, and is just as anxious to admit the limits of his understanding – by declaring himself to be in the conversation for the long haul. He will not, if his definition is refuted, sulk – as a Meno, say (*Meno* 80d) might have been tempted to do. He is aware that inquiry is difficult, but he is willing to do his bit: to reflect and do his best to form a belief with the sort of shape Socrates needs it to have.⁷¹ He is willing to advance it, *as his belief*, without reluctance or qualification, even while forswearing any dogmatism with respect to it. It is in this last combination of attitudes that I will now argue the key to our problem is to be found.

⁷¹ The most thorough recent account of Socrates' expectations for a definition is Vasilis Politis, "What Is behind the Ti Esti Question?," in *The Development of Dialectic from Plato to Aristotle*, ed. Jakob L. Fink (Cambridge: Cambridge University Press, 2012), 199–223.

I began with a worry about how we are to interpret Socrates' demand that his interlocutors say only what they believe: what picture of self-knowledge, and what picture of belief, does the demand presuppose? The assemblage of passages above has begun to suggest what that picture might be, and I shall now proceed to argue that it is a richer, more interesting, picture than any previous readings of these passages have managed to capture.

5. Theoretical or Deliberative?

The passages from the *Theaetetus* and *Philebus* quoted in the last section give us, as we have seen, the materials for a theory that connects thought, belief, assertion and self-knowledge. I have suggested that this model may be fruitfully used to analyse the moments in the dialogues where Socrates asks his interlocutors to assert or assent to some belief. But not all such moments seem to cry out for analysis. Indeed, the dialogues are rife with cases of interlocutors – naïvely, thoughtlessly, unselfconsciously – complying with Socrates' demand that they say what they believe without his having to lay down the law. It is, as we have seen, the cases where he *does* need to lay it down that call for some explanation, in terms of his

interlocutors' motivations for saying something other than what they believe.⁷² This must have something to do with the fact that in the very simplest cases, assertion and assent are straightforward affairs. The relevant dispositions of the interlocutor being simple and determinate, all they have to do in order to say what they believe is simply to come out with the first thing they are inclined to say.⁷³ Unless, obscurely enough, something has 'gone wrong', no more effort than this will be necessary to comply with the Rule.

The cases that do call for analysis in terms of some model of belief formation are those – discussed above – where Socrates' interlocutors show signs of hesitation, reluctance and effort. Correspondingly, Socrates reacts to their assertions with praise, as well he might to the relative advance in self-knowledge that those assertions so often represent. I have suggested that the effort in question is that of articulating a complex disposition in words; the more complex the disposition, the

⁷² A helpful discussion of this general phenomenon appears in Richard Moran, "Problems of Sincerity," *Proceedings of the Aristotelian Society* (Hardback) 105, no. 3 (June 1, 2005): 348: 'expressing something other than one's actual beliefs is something that calls for some explanation, specifically an explanation in terms of the motives of the speaker regarding his audience'.

⁷³ See Bernard Williams, *Truth and Truthfulness: An Essay in Genealogy* (Princeton: Princeton University Press, 2002, 76: 'It follows from the basic spontaneity of assertion that [the disposition of] Sincerity does not typically involve a special exercise ... in discovering what it is that I believe; rather in the simplest case I am confronted with my belief as what I would spontaneously assert.' Similarly, Moran, 'Sincerity', 348: 'the natural, spontaneous response to any question calling for an assertion is to come out with one's actual beliefs'.

more of an achievement it is to articulate it adequately. But there is more than one way of thinking about what this effort at articulation involves.

Modern philosophy of mind – a tradition that extends from Descartes to the writings of contemporary analytic philosophers and their counterparts in the ‘Continental’ phenomenological tradition – encompasses a wide range of views on this question. The broad structure of the modern debate has to do with the analogies and disanalogies between our knowledge of our own minds and our knowledge of the minds of others; the question of belief is, in these debates, a special case of a more general question about all mental states, both attitudes and sensations. At one end of the spectrum is a set of views – sometimes labelled ‘Cartesian’ – that claim a strong disanalogy between the first- and third-personal cases.⁷⁴ One thing that unites these views is their taking seriously a *perceptual* model of self-knowledge, where one comes to have knowledge of (say) one’s beliefs by taking them in with a metaphorical ‘inward glance’. At the other end of the theoretical spectrum is a set of views – call them ‘deflationary’ – that deny that the disanalogy between the first- and third-personal is anywhere as deep as the Cartesians claim. The different claims pertain to a variety of micro-questions: is our

⁷⁴ It is, as is so often the case with such labels, a real question whether Descartes himself was a Cartesian in this sense. For a brief discussion and bibliography, see Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001), 12 n. 9. The inverted commas around Cartesian should be assumed in what follows.

knowledge of our mental states infallible? Is it 'immediate', i.e. non-inferential? Is it helpfully likened to perception? Is one's access to the contents of one's mental states somehow *privileged*? Are there good grounds for the widespread practice of treating agents, by default, as authoritative about their mental states?

That last question – the problem of first-person authority – is the most important for our present purposes. No one denies that there is more than one way to know someone else's beliefs: observation, interpretation, and forth. But there does seem to exist a default convention of deference to self-ascriptions of beliefs. If someone says 'I believe that p', then there is – on the face of it – something improper and presumptuous about replying 'No you don't.' Rivals theories have different explanations of this convention, with explanations ranging from the metaphysically ambitious (e.g. first-personal introspective self-ascription is infallible while third-personal interpretative conjectures are not) to the deflationary (e.g. it is merely a social convention that we defer to people's avowals of belief).⁷⁵ But the fact of the practice itself is common ground to all sides of the debate.

⁷⁵ Something like the latter position is argued in Crispin Wright, "Self-Knowledge: The Wittgensteinian Legacy," in *Knowing Our Own Minds*, ed. Crispin Wright, Barry C. Smith, and Cynthia Macdonald (Oxford: Clarendon Press, 1998).

On the face of it, Plato's Socrates is no exception here: he too seems to treat self-ascriptions of belief as having an authority that third-personal interpretative claims do not. Why else is he so insistent on asking people to say what they believe? But this is evidently too quick. For one thing, it is not in fact clear that Socrates consistently grants self-ascriptions of all mental states that authority. Most famously, he operates on the default presumption that most if not all self-ascriptions of *knowledge* are unfounded. Much of his practice is concerned to undermine just this sort of self-ascription. But of course, knowledge – a 'factive' state as we should now say – is a special case: a claim to knowledge can be undermined if the proposition claimed to be known is false. Equally, there are ways of getting someone making a knowledge claim to offer some demonstration of competence in the relevant subject matter; the ability to offer such a demonstration at least evidential, perhaps even criterial, of knowledge.

Belief is not factive in that way. But there are still similar ways of undermining a claim to believe something. Professions of belief can certainly be undermined by behavioural evidence; consider the ways in which hypocrisy is exposed. But the assumption of first-person authority was only supposed to be a *default* attitude. Extraordinary circumstances – dissembling, repression – call for a change in the default attitude. It would be extraordinary for a judge, or a psychoanalyst, to take

someone's self-ascription of belief to be decisive. A self-ascription is in those circumstances to be treated with some degree of scepticism, a scepticism born of the judge or analyst's awareness of why the witness or analysand's utterances may not, however loud their protestations, express their real beliefs. Socrates is alert to some of these possibilities. Consider his remarks to Glaucon and Adeimantus in the *Republic*, when he has just heard them speak on behalf of injustice while professing not to believe what they are saying ('It isn't, Socrates, that I believe any of that myself. I'm perplexed, indeed, and my ears are deafened listening to Thrasymachus and countless others', 358c):

... you must indeed be affected by the divine if you're not convinced that injustice is better than justice and yet can speak on its behalf as you have done. And I believe that you really are unconvinced (*alēthōs ou pepeisthai*) by your own words. I infer (*tekmaïromai*) this from the way you live (*tropou*), for if I had only your words (*logous*) to go on, I wouldn't trust (*ēpistoun*) you. (*Republic* 368ab)

Socrates has three pieces of evidence to go on: an earnest avowal (*Republic* 358c) that they do not believe the Thrasymachean doctrine they are about to defend, their impassioned defence of injustice and the conviction they are able to affect while delivering it, and the kind of life he knows they lead. It is the third piece of evidence that clinches it. Faced with the cogency of their case for the other side, it is only his knowledge of how they live that convinces him they do not really believe what they are saying. Of course, it would be at least strange if Glaucon or Adeimantus were to make the same claim: 'Given the cogency with which I have

spoken in favour of injustice, it is only an observation of my behaviour that gives me grounds for claiming that I do not really believe what I said then.' I shall have more to say about this strangeness, and the passage itself, in due course, but it seems to me to show that Socrates is perfectly willing to disregard first-personal authority when necessary, and rightly so.⁷⁶

A further point is relevant here, and its classic articulation appears in these much-discussed remarks of Gareth Evans':

... in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward – upon the world. If someone asks me "Do you think there is going to be a third world war?," I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?"⁷⁷

I take the idea to be that when one is asked a question about one's (present) beliefs, one answers it by considering the very same reasons that would justify one's answer to the corresponding question about the world. But this seems to elide a distinction we might want to make, between discovering a belief and forming one, between a procedure that reveals beliefs and one that creates them. But how important it is to maintain such a distinction, and how sharply it should be drawn, is a further question. Plato's dialogues give us some material for reflection on just this point. Let

⁷⁶ In this, I am largely in agreement with the central argument of Woolf, "Socratic Authority".

⁷⁷ Gareth Evans, *The Varieties of Reference*, ed. John McDowell (Oxford: Oxford University Press, 1982), 225.

us return to the *Charmides* passage considered earlier. If Socrates' injunction to Charmides is taken (more or less) literally, then he does seem to assume something like the 'inward glance' idea we have been calling Cartesian:

... look into (or at) yourself, with greater concentration, and when you have decided what effect the presence of temperance has upon you and what sort of thing it must be to have this effect, then put all this together and tell me clearly and bravely, what does it appear to you to be? (*Charmides* 160de)

But this is precisely what Charmides does not do.

He paused (*epischōn*) and, looking into himself (*diaskepsamenos*) very manfully, said, "Well, temperance seems to me (*dokei ... moi*) to make people ashamed and bashful, and so [...] modesty must be what temperance really is." (*Charmides* 160e)

Despite what Socrates says – and how ever could he know? – there is no reason to suppose Charmides is looking into or at *himself* at all. He seems – and his words provide the evidence we need – rather to be looking 'outward', at the fact that temperance tends 'to make people ashamed and bashful', and he uses this (not inaccurate) observation as the grounds for an inference to a generalisation of precisely the sort Socrates wants in a definition. The natural way to interpret what Charmides has said is that he has just *formed* a belief based on evidence derived from experience. Does this mean Charmides has disregarded Socrates' instruction? In one, trivial, sense, he has, by looking towards something other than himself. But Socrates does not seem to care. It seems, on the contrary, that Charmides has done what Socrates usually wants in these situations: articulate a belief from what materials are at hand. It is certainly what he seems to want from Theaetetus, where

the metaphors of perception are absent from Socrates' instructions to him. (It is perhaps no coincidence that the subsequent stretch of the dialogue is an inquiry into perception; perceptual metaphors cannot have struck him as innocent while he was composing those passages.) The language in that passage is, rather, one of confidence and authority, language that is also there, if more mutedly, in the *Charmides*. Having given Theaetetus, with the aid of a toy geometrical example, an idea of what he expects from a definition (a familiar Socratic manoeuvre), he now says:

... do have confidence (*tharrei*) in yourself ... You must put your whole heart (*prothumēthēti*) into what we are doing ...

If putting one's heart into it (*prothumias*) is all that is required, Socrates, the answer will come to light (*phaneitai*).

Go on, then. You gave us a good lead just now. Try to imitate your answer about the powers. There you brought together (*perielabes*) the many powers within a single form; now I want you in the same way to give (*proseipein*) one single account of the many branches of knowledge. (*Theaetetus* 148cd)⁷⁸

A shift in vocabulary is discernible, from the quasi-introspective language of the *Charmides* – which is in any case belied by how Charmides actually interprets it – to the more 'agential' tone of the *Theaetetus*.⁷⁹ But what does this shift amount to, if

⁷⁸ See also *Theaetetus* 187a–c discussed in the last section, and Theaetetus's later reference to this exchange at 204b: 'I don't feel at all certain; but as you *keep telling me to answer with a good will* (*prothumōs*), I will *take a risk* (*parakinduneuōn*) and say they [sc. 'sum' and 'whole'] are different.' Socrates replies: 'Your good will (*prothumia*), Theaetetus, is all that it should be. Now we must see if your answer is too.' To which Theaetetus, ever the good sport, replies: 'We must, of course.'

⁷⁹ Some readers might detect in this shift another Platonic change of mind. But given that the introspective – or vulgar Cartesian – picture is not ever stated or argued for in the early dialogues,

anything? We can get some hold on the question if we consider how Socrates' question – effectively, 'What do you believe?' – might sound to his interlocutors when they put it to themselves in a first-personal form. The question 'What do I believe?' is in fact ambiguous between two construals that we might label the 'theoretical' and the 'deliberative'. In its theoretical sense, it may be rephrased: 'What is it that I believe?' and in its deliberative sense, 'What am I to believe?' The first question is answered in a discovery, the second in a resolution.⁸⁰

Discovering that one believes something is a perfectly ordinary occurrence, and it is not always an achievement. When our dispositions – of assertion, assent and behaviour – are sufficiently determinate, a spontaneous, uninhibited, assertion is all it takes to articulate it. There is some sense in which the belief expressed with a spontaneous assertion has only just then been formed; but insofar as the

the fact of a suggestive shift in vocabulary is too weak a basis for a developmental claim. It is of course possible that Plato came to have reservations about the introspective language of the earlier dialogues and consciously chose more deliberative metaphors when composing the *Theaetetus*, but this is rather less than a change of mind. The best case for treating Socratic dialectic in the *Theaetetus* as continuous with that in the early dialogues is M. F. Burnyeat, "Examples in Epistemology: Socrates, Theaetetus and G. E. Moore," *Philosophy* 52, no. 202 (October 1977): 381–98. Burnyeat claims, to my mind plausibly, that 'Socrates' dialectical method ... is recognizably practised in the *Theaetetus*, albeit on a larger scale and with a content that reflects Plato's own later concerns, including a concern with methodology, not the views of the historical Socrates' (381 n. 1). In this, the *Theaetetus* is importantly unlike either the *Republic* on the one hand, or the *Sophist* and *Statesman* on the other, though these dialogues all have in common the fact that they are long and constructive. On this theme, see also the suggestive remarks in Sedley, *Midwife*, 13–15 and *passim*.

⁸⁰ The terminology is from Moran, *Authority*, 56 ff. The contrast is most succinctly stated as 'the difference between that inquiry which terminates in a true description of my state, and one which terminates in the formation or endorsement of an attitude' (63).

dispositions it articulates have a longer prehistory, it is a perfectly ordinary use of the term belief to say one has always had that belief.

Of course, our dispositions are not always so determinate, and that is why we sometimes demur when asked to state our beliefs: we are not sure what they are, or even if we have any. The point may be pressed with relation to Socratic practice, for Socrates' questioning rarely brooks demurrals. Indeed, this insistence on decisive self-ascription may well be held against him.⁸¹ Why the pressure to declare a belief? Why does Socrates not allow his interlocutors the right to their uncertainty? Why does he not distinguish between asking his interlocutors to express a long-held – and pre-articulated – opinion and asking them to articulate one from a jumble of dispositions in response to his questioning?⁸²

⁸¹ The case for the prosecution is put forcefully in Beversluis, *Cross-examining Socrates*, 46: 'Socrates seldom acknowledges his interlocutors' reluctance to give their unqualified assent. Indeed, he habitually minimizes (or ignores) their misgivings and presses for their assent ... These psychologically manipulative tactics suggest that, although officially insisting on sincere assent, he is often willing to settle for qualified, and sometimes heavily qualified, assent.'

⁸² The point is, again, sharply pressed in Beversluis, *Cross-examining Socrates*, 50: 'Since it is intelligible to say that Socrates' interlocutors held beliefs before their dialectical encounters with him, it is important to determine whether the beliefs they bring with them into the discussion are the same as the propositions they assert (or to which they assent) during it. If there are good reasons for thinking that they are not, then there are also good reasons for denying that Socrates elicits their sincere assent.' A great deal rests on the idea that his interlocutors 'bring [beliefs] with them into the discussion': but given what I have been arguing, we can reasonably ask in what sense they do this. Given Beversluis's own, apparently dispositional, understanding of beliefs, what they bring into the conversation is a set of discursive and behavioural dispositions, not all of which need be linguistically inflected. But unless he claims, implausibly, that there is no pattern to their assertions in their exchanges with Socrates and they are answering arbitrarily, those assertions must be articulations of at least *some* of their pre-existing dispositions. Beversluis must be saying then that

If the question ‘What do you believe?’ is interpreted as a theoretical question, then it might be entirely correct to answer ‘I don’t know’ or ‘I’m not sure’. But things look different when the question is interpreted as a deliberative demand. To demur in response to a deliberative demand is really a refusal of sorts: a refusal to make up one’s mind, or a refusal to speak one’s mind even though one’s mind is made up. Neither is necessarily objectionable, and it is possible to think of any number of situations where it is a quite proper response. But neither ignorance nor uncertainty could by themselves justify such a response, for an invitation to adopt (what we might call) a deliberative stance is an invitation that says, ‘If you don’t know, *find out*’ and ‘if you’re not sure, *decide*’.⁸³

Socrates is likely aware that his interlocutors have complex dispositions relevant to the subject matter he discusses with them. He is, after all, constantly being

the articulation is somehow incomplete: the hesitation of the interlocutor is a sign that he also has other, contradictory, dispositions that have not been expressed in the proposition to which he has assented. With this I agree, but I submit that contrasting an interlocutor’s pre-existing beliefs with his dialogic assertions begs the question. Why treat only the pre-existing dispositions as deserving the name belief and not those manifested in their responses to Socrates?

⁸³ The use of the ‘practical’ vocabulary of decision and deliberation to apply to thought has precedents even in archaic Greek discourse. See Bernard Williams, *Shame and Necessity* (Berkeley: University of California Press, 1993), 36: ‘... Homer has no word that means simply “practical deliberation”. ... But this is equally true of modern English, as when we speak of “wondering” or “considering” what to do; even the word “decide” itself is not confined to practical connections. ... Language itself, our own and Homer’s, can help to remind us that decision is not a special kind of action, and also that wondering what to do has something in common with wondering how things are: they are both fully resolved by one’s becoming certain.’

presented with evidence of their competence at *ostensive* definition.⁸⁴ But while he is completely consistent in pronouncing such definitions inadequate as answers to his ‘What is X?’ questions, he does not – as some have alleged – pronounce their ostensive talents irrelevant.⁸⁵ On the contrary, he is often asking them to ‘put it all together’, i.e. perform some effort of synthesis, and then to give voice to the resolution they have achieved in an assertion.

To summarise: I have suggested that there are two ways in which a Socratic interlocutor might interpret a call to ‘say what he believes’: theoretical and deliberative. I am now suggesting that Socrates, in asking his interlocutors to say what they believe is *asking them to adopt the deliberative stance*: not to cast around ‘in’ their minds looking for any old belief-shaped thing to bring to class for show-and-tell, but to exercise responsibility for bringing some sort of order and stability to their (likely chaotic) behavioural and discursive dispositions, and to affirm it as theirs. This seems to me the best explanation for the preponderance of such vocabulary in the texts I have been discussing.

⁸⁴ E.g. *Euthyphro* 5d, *Meno* 71e–72a, *Theaetetus* 146cd. Charmides, to his credit, unites ostension with generalisation on his first try (*Charmides* 159b)

⁸⁵ E.g. P. T. Geach, “Plato’s *Euthyphro*,” *The Monist* 50, no. 3 (1966): 369–82. See also the discussion of Geach’s charges in Burnyeat, “Examples”, 382 and *passim*.

Even when interpreted this way, we can press the traditional objection: how fair a demand is this? Can it ever be proper to require people to adopt such a stance with respect to their beliefs – or, as we might now say, their proto-beliefs? There are certainly situations in which the demand can be presumptuous, or unnecessary, or even intellectually injurious.⁸⁶ Unless there are special circumstances, no one owes anyone else a resolution of their doxastic ambivalence. Further, not every belief even needs to be articulated for it to play its part in the economy of the mind and the conduct of life; the unarticulated belief is certainly worth having. Further, even if there is something to be said for articulation, there are dangers to premature articulation, forcing oneself to put one’s weight behind some formulation of one’s beliefs without giving the matter the time and attention it deserves.⁸⁷ The splendid ‘maieutic’ metaphor of the *Theaetetus* – Socrates as a midwife of ideas – implies this: ideas, like babies, need time to gestate. A prematurely articulated idea, like a prematurely born child, is vulnerable and dependent as one carried to term is not, or not to the same degree. If the analogy runs further, then a questioner who demands a premature articulation of an idea is just as irresponsible as a midwife who coaxes a baby out prematurely. Socrates claims that ‘the most important thing

⁸⁶ A classic statement of the anti-Socratic case is Geach, “*Euthyphro*”.

⁸⁷ The virtues of patience in this respect are eloquently put in a remark of Benjamin’s; Walter Benjamin, *Selected Writings: 1913-1926* (Cambridge, Mass.: Harvard University Press, 1996), 458: ‘The more circumspectly you delay writing down an idea, the more maturely developed it will be on surrendering itself.’

about my art is the ability to apply all possible tests to the offspring, to determine whether the young mind is being delivered of a phantom' (150c) – but this is a stranger remark than is usually acknowledged. Can Socrates – who now stands charged with intellectual infanticide in addition to everything else – be defended?

6. Sincerity, Willingness and Socratic Protreptic

The case against Socrates rests on the existence of situations where his methods would be impertinent, fruitless or pernicious.⁸⁸ Any defence of his methods must therefore be grounded in an account of the situations where they would be proper, fruitful and beneficial, and an argument to the effect that (at least) the encounters depicted in Plato's dialogues belong to this latter category. These situations would need to be of a sort where someone with no articulate belief on some question may have reason to form one on the basis of a set of determinate dispositions he has, or where someone with competing dispositions amenable to multiple distinct and incompatible articulations has reason to attempt a resolution in the form of a single articulation. There is also the further question of whether these articulations should be made public, and whether Socrates is a suitable confidant for these articulations.

⁸⁸ Socrates comes close to admitting this at *Theaetetus* 150e, very possibly an allusion to the well-known case of Alcibiades (also alluded to in the *Republic* and depicted richly in the *Symposium*), but attributes the ill effects to the young men in question leaving him 'sooner than they should, either of their own accord or through the influence of others.'

We can speak of reasons to form or not to form a belief (i.e. to form a belief at all, as opposed to forming one belief rather than another) because we recognise the notion of an ‘investigative investment’, i.e. that articulating our dispositions into beliefs takes time and effort that could well be spent on other things.⁸⁹ It is for this reason that it stands in need of justification, justification that is more easily provided in some cases than in others. Sometimes, the investment is partially constitutive of some role, say that of doctor or scientist or police detective. Sometimes, curiosity is justification enough. Sometimes, some practical need – the need to offer advice or testimony to someone to whom one feels they are owed – justifies the investment. Similarly, there are cases where to keep one’s beliefs private would be objectionable. Again, we may mention the doctor or lawyer asked by a patient or client for a medical or legal opinion, or a politician canvassing for votes asked for her views on questions of public policy.

Certainly, one ought not to demand a *premature* articulation, but this is a trivial point. The real question, for which there is likely no general answer, is when it is proper to make such a demand. But at least in the professional examples above, it is

⁸⁹ The term in inverted commas is from Bernard Williams, who defines it as ‘the idea that information, acquiring a true belief about a given question, can have a cost, in time, energy, opportunities lost, perhaps dangers run. So there are questions such as “How much trouble is it worth to find out about this?”’ (Williams, *Truth and Truthfulness*, 87).

clear enough that there are some built-in limits on just how long one may properly defer articulation. A responsible doctor cannot endlessly defer offering a diagnosis, and a politician facing the electorate cannot defer expressing an opinion on policy. Anyone purporting to cure disease or exercise political power has, in so purporting, legitimised certain discursive demands. To make such a demand in these cases remains reasonable when the task is difficult and the requisite investigative investment correspondingly high. To ask for a doctor's considered medical opinion on a complex case is not to assume that complete certainty is possible; on the contrary, to do so is to ask: 'Allowing that complete certainty is impossible, behind which of several possibilities are you willing to put your weight?'

The cases more relevant for our purposes are those where one might reasonably be reluctant to make a *knowledge* claim because one is ambivalent or uncertain.

Philosophers are well aware of such cases. But our question here pertains to belief. Certainly one should not claim to know if one does not know, especially given that on most questions of importance, it is not wholly in one's power simply to 'find out' the correct answer. But there are parallel claims with beliefs, particularly the sorts of beliefs we have been considering (medical or legal opinions) which are both authoritative and consequential. In giving one's professional opinion, even when one cautions that one might very well be wrong, one is nevertheless giving other

people a reason to rely on one's words. No doubt the costs are different, and lower, from those cases when one makes a knowledge claim. Still, the relevant notions of responsibility, and the idea that one can quite properly be held to account for one's professional opinions, remains the same.

A great deal here turns on the some idea of what is *at stake* when someone expresses a belief. Much of life consists of people expressing largely inconsequential beliefs.

Here, some degree of irresponsibility, eccentricity, carelessness is forgivable. There is much more at stake in the question 'Is chemotherapy in your professional opinion likely to do much good?' than in 'Is it going to rain, do you think?', though the latter question imposes different responsibilities on the professional meteorologist than it does on a layperson who has claimed no special talent at predicting the weather.

But if there is only a duty to take on some investigative investment rests because of the likelihood that other people will rely on our authoritative pronouncements, then this does not seem to describe the situation of most of Socrates' interlocutors in their conversations with Socrates. Socrates is not entitled to their opinions in virtue of any role-based relationship he has with them, and he rarely gives any impression of relying on their pronouncements in such a way as ought to make them keenly aware of their epistemic responsibilities to him. The closest thing to an

exception is Euthyphro, whom Socrates meets outside the king-archon's courts where he (Socrates) is shortly to be charged by a certain Meletus with impiety and the corruption of the youth. Euthyphro, who is there to prosecute his father for murder, is contemptuous of his father and other relatives, who think his prosecution impious: 'their ideas of the divine attitude to piety and impiety are wrong' (*Euthyphro* 4e). Socrates replies with his characteristic irony to the overconfident assertion: 'It is indeed most important, my admirable Euthyphro, that I should become your pupil' (5e) and says he intends to defer to his (Euthyphro's) superior wisdom on matters of piety in conducting his own defence. We have seen enough of Euthyphro even by this point in the dialogue to recognise that Socrates cannot be serious, but there is no evidence that Euthyphro is sensitive to this. Socrates is inviting Euthyphro to share his claimed wisdom with him, with all the sense of responsibility involved in offering it as – in effect – legal advice to a defendant.

In fact, this case is anomalous. On the face of it, if Socrates' interlocutors have responsibilities, they are not *to him*. He is not obviously harmed if they fail to say what they believe, or fail to put enough effort into making up their minds. This is the sense in which Socratic method is often described as therapeutic: it is good for the patient, but not (necessarily) for the therapist. In the *Charmides*, this thought

itself is put to some use: Charmides is enticed into conversation with Socrates with the promise that conversing with him might cure him of his recurring headaches. If he believes this, then it would simply be beside the point for him to dissemble, just as it would be to dissemble a doctor about (what one can say of) one's symptoms.⁹⁰

There is a further sense in which some of Socrates' interlocutors do have responsibilities to him, and this is in their capacity (not always taken on voluntarily) as fellow inquirers. If two people regard themselves as engaged in an cooperative inquiry, it is at least strange if one of them is reticent about revealing relevant beliefs of his, and it reflects poorly on his involvement in the inquiry if he refuses to do the necessary work to articulate a belief or make a judgment once one is solicited by one's fellow inquirer.⁹¹ Socrates frequently uses such language.⁹² The modern analogy usually invoked in this connection is psychoanalytic practice.

⁹⁰ A thorough development of the medical analogy, in the context of a scholarly discussion of the Hellenistic philosophical tradition, is Martha C. Nussbaum, *The Therapy of Desire: Theory and Practice in Hellenistic Ethics* (Princeton: Princeton University Press, 1996), 16–37 and *passim*.

⁹¹ The point is well put in Alasdair MacIntyre, "Truthfulness, Lies, and Moral Philosophers: What Can We Learn from Mill and Kant?," *The Tanner Lectures on Human Values* 16 (1995): 333: "Truth is the good internal to rational enquiry and the kind of trustworthiness required from each other by those who participate in enquiry includes an unfailing regard for truth and for truthfulness."

⁹² E.g. *Meno* 80d: 'I want to examine and seek together with you what it may be'; *Charmides* 158e: 'I would like to investigate the question with you'; also 166cd. See also *Laches* 184e ff where the verbs of investigation (e.g. *zētoumen*, *skopoumen*) are generally in the first-person plural, *Gorgias* 505e–506a and *Protagoras* 348c–e. Among the later dialogues, see *Theaetetus* 150a–151e, *Sophist* 218b5–d6, *Statesman* 258b–c, 285c–d and *Philebus* 19a–c, especially c4–6. The most conspicuous association of doxastic vocabulary with deliberative in Plato appears in one of Socrates' etymological conjectures in the *Cratylus*, at 420bc, but the tone there is, to my mind, too studiously playful for any substantive interpretative thesis to gain much by appealing to it.

Indeed, that tradition is *nonpareil* in its attention to the distinctive difficulties involved in saying what one believes, difficulties explained by a certain approach to explaining our all-too-human failures of self-knowledge in such phenomena as repression, introjection, and transference.⁹³ I shall try to explain the point using the terms of a less exotic (to me) setting that has also been compared to the practice of Plato's Socrates: the Oxbridge philosophy tutorial.⁹⁴ (The comparison is natural: after all the conduct of the tutorial has long been informed by interpretations of Socratic practice.)

There are, of course, as many approaches to the conduct of a philosophy tutorial as there are tutors, but there are common features more or less constitutive of the practice. The student is assigned an essay, based on at most a week's worth of reading and thinking and a question or 'title' to focus the mind. The essay is read out in tutorial or handed in in advance and becomes the focus of the subsequent hour or so of discussion with a tutor and often a student tutorial partner who has

⁹³ The best brief treatment of the connection between Socratic and psychoanalytic practice is Jonathan Lear, "The Socratic Method and Psychoanalysis," in *A Companion to Socrates*, ed. Sara Ahbel-Rappe and Rachana Kamtekar (Oxford: Blackwell Publishing Ltd, 2007), 442–62. Lear is illuminatingly attentive to the extent of the putative disanalogy between the centrality of beliefs to the success of Socratic practice and the relative unimportance of beliefs to conceptions of psychoanalytic success. Similar questions are addressed in connection to Aristotle and the Hellenistic philosophers in Nussbaum, *Therapy*, *passim*.

⁹⁴ The contributions by Robin Lane Fox and Alan Ryan to David Palfreyman, ed., *The Oxford Tutorial: Thanks, You Taught Me How to Think* (Oxford: The Oxford Centre for Higher Education Policy Studies, 2008) have instructive discussions of the aptness of the analogy.

also been set the same work. The tutor's remarks are frequently in the interrogative mood. Requests for clarification and restatement abound ('When you said... what did you mean exactly?') Distinctions are urged, terminology defined and defended. Objections are presented, sometimes the tutor's own, sometimes with the tutor taking on for pedagogical purposes the point of view of some philosopher whose view has been challenged in the student's essay. Counterexamples are presented, and attempts are made either to resist the force of the counterexample, to argue for its irrelevance, or to modify the initial claim so as to make it invulnerable to it.

There are at least two slight disanalogies with Socratic practice. For one thing, the undergraduate is no ingenuous youth thrust into conversation (though the Socratic interlocutor is rarely, despite the cliché, all that ingenuous either); she – who unlike the Socratic interlocutor can (now) in fact be a she – is there willingly, and has been given materials with which to prepare for the encounter. For another, the tutor rarely professes to know nothing about the subject being discussed, though the questions set for essays will tend to be those on which there is as yet no academic consensus and most tutors will avoid using the setting as an opportunity merely to evangelise for their own views and which most will have the grace to designate as views.

An important analogy between the two practices is that the tutorial, like the Socratic elenchus, is initiated by a *question* on one side, and a *submission* on the other. A undergraduate essay is unlike a interlocutor's definition insofar as a student has had a week to think about the question, and any amount of assistance in the way of reading lists and introductory surveys. They are certainly not required to introspect their way to an answer – and it is not in any case clear how one could introspect one's way to an answer to such a question as 'Are Plato's arguments for the tripartition of the soul sound?' This does not mean nothing akin to introspection is involved in answering such a question. The undergraduate, at any rate if she is conscientious, will evaluate some of the arguments with which the assigned reading confronts her by asking herself how plausible she finds their premises, and where she finds them implausible, will attempt to articulate why. The notion of plausibility – and its more demanding sibling 'intuitiveness' – are essential to the practice of reflective reading, and of philosophical reflection more generally. To ask 'Is p plausible?' is more or less equivalent to 'Should I believe that p?'

Like Socrates' interlocutors, an undergraduate is required to do more than offer a competent survey of the theoretical alternatives in 'the literature' (the final two sections of this essay will have more to say about the classical analogues of this modern academic category). She is required to express an opinion and defend it.

She may well feel ambivalent about the essay she hands in or reads out – quite apart from the fact that she has not had much time to write it, the questions set are rarely such as to be ‘answered’ in the 1,500 or so words she has. But this fact does not change the attitude of the tutor to it, an attitude that can often be mistaken for antagonism or hostility. In the same way that a chess player has no obligation not to take his opponent’s queen simply because the opponent was evidently ambivalent about his decision to leave her undefended, the tutor is not required to ‘go soft’ on an unsound argument.

The aims of the tutorial are governed by a host of norms that stand in a range of subtly different relations to the notion of truth: belief, knowledge, accuracy, relevance, honesty, rigour, consistency, coherence, plausibility, etc. But the distinctive form of their application in the tutorial can only be understood in relation to the two speech acts that initiate the tutorial: the asking of a question and the making of a claim. A tutor might well say, à la Socrates, that the student must say only what she believes. But the philosophy tutorial is the paradigm case of a setting where this way of putting the demand risks absurdity. What student has well-articulated views on (say) eliminative materialism or the private language argument or the theory of forms before she sets about getting through her weekly reading? It is of course possible to give some sense to the demand: one might wish

to urge the student not to defend a position simply because she takes it to be the one her tutor holds, or simply the one easier to defend, or the only one she took the trouble to read about. And on those questions – say, the existence of God – on which many students may well hold opinions one way or the other, it might be valuable to make it clear that nothing in the nature of philosophy or the institution of the tutorial disallows the expression of their actual opinions. The point is, of course, not the claims themselves but the arguments given for them. And one thing a student does in asserting and arguing a position is tacitly consent to having that position taken up as the subject of a discussion of certain sort.

Not all assertions imply such consent;⁹⁵ even when they do, one's interlocutors may wish to waive their newly conferred rights to challenge the assertion. To do so might, for instance, be impolite, or bothersome, or simply tedious. But the tutorial is a special setting. The position defended in a student's essay is not some set of casual remarks, and the tutor would be failing to do her job if she treated them as if they were. When a tutorial fails as a result of a student's attitude, it can either be because she shows a general reluctance to take a view at all, or because she clings to every view she has expressed with a degree of commitment that precludes any

⁹⁵ This is one of many reasons to hold back from accepting in any unreconstructed form the 'inferentialist' picture of 'discursive practice' in general, such as the one defended in Robert Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Cambridge, Mass.: Harvard University Press, 1998), on which I am of course drawing heavily in this section.

attempt on the tutor's part to bring rational considerations to bear on them. The first – call it wishy-washiness – is a problem in this context because it shows a disregard for the truth-seeking aims of the tutorial; the tutorial aims to uncover truth, or at any rate to reduce falsehood, and to refuse to endorse any view at all is to refuse to join in the game, to enter into the practice. The second – call it dogmatism – shows a similar disregard for the intimate relationship between belief, rationality and truth: sometimes one may well want to stick to a view that has, by all accounts, been refuted, but to do that would be to play a different game. Either way, something has gone wrong.

A further point may be made here: the tutorial, despite the difference in the official intellectual credentials of its participants, is in fact a species of joint inquiry.⁹⁶ The aims of the two participants are, ultimately, the same – truth and knowledge – though it may well be that one participant has more to offer the other than *vice versa*. In some successful tutorials, however, the conversation will approximate one between peers, at least to the extent that the tutor will come away having learnt as much as the student. The feature of a tutorial that makes it a kind of joint inquiry is the fact that it is an encounter into which one enters well aware that one's beliefs

⁹⁶ This fact is reflected in the traditional, and increasingly old-fashioned, locution used to describe the encounter: 'Lucy is *reading Plato with* Professor Snodgrass'. The image it evokes of tutor and student reading in a companionable silence is amusing, but it captures an important truth about the encounter.

will be subject to challenge, and therefore, potentially to being changed. The tutor, while better armoured against this possibility, is hardly invulnerable, and no good tutor would want to be.

The thought most directly pertinent to our discussion of Socrates is this: it is essential to certain kinds of discursive practice that the participants in that practice enter into it willing to subject themselves to scrutiny. In order for this to be possible, they must be willing to articulate and take ownership of a claim that they are either uniquely inclined to endorse (as will sometimes be the case), or one that they are more inclined to endorse than any alternative that occurs to them (likely the more common occurrence). In asserting something in such a setting, one makes a *claim*, and to make a claim is to invite assessment in terms, among other things, of what other claims one is inclined to make and the extent to which one can rationally claim all those things at the same time. The aim of such a practice is, then, to continue in public a process one has already begun in private – if only as recently as the question was first made salient to one’s mind (and if only by the thought of an imminent deadline). In deciding to assert one thing rather than another, one has reached a preliminary resolution, but it would be a kind of intellectual vice to assume no further work will be necessary.

I have argued in this section that Socratic practice is a discursive practice of roughly the sort described above. If this is right, then for Socrates to require of his interlocutors that they say (only) what they believe in such a context does not rest on the implausible assumption that his interlocutors enter the conversation with pre-articulated beliefs. Rather, it is to urge them to do their best at articulating what content is in their minds to the point that it comes to have some of the characteristic features of a belief, rather than (say) a supposition or a fancy. These features include a willingness to endorse some linguistically articulated content, and a further willingness to take responsibility for that content by tacitly licensing a certain kind of evaluation of that content in terms (among other things) of its consistency with whatever other content in their minds they are prepared to give a similar endorsement to.

Readers inclined to worry about a familiar feature of Socratic practice – viz., his apparent disregard for his interlocutor’s evident ambivalence about what they are asserting or assenting to⁹⁷ – might well find some reassurance about Socrates’ good

⁹⁷ E.g. Beversluis, *Cross-examining Socrates*, 46 puts the point forcefully: ‘Socrates seldom acknowledges his interlocutors’ reluctance to give their unqualified assent. Indeed, he habitually minimizes (or ignores) their misgivings and presses for their assent: by asking leading questions, by resorting to mock flattery, by employing shame tactics, by soliciting the aid of other auditors, and even by threatening to leave These psychologically manipulative tactics suggest that, although officially insisting on sincere assent, he is often willing to settle for qualified, and sometimes heavily qualified, assent.’ As Christopher Gill, “Speaking Up for Plato’s Interlocutors. A Discussion of J.

faith in this fact. While there is an important difference between the sort of enthusiastic assent expressed in a ‘*panu ge!*’ (‘Certainly!’) and one expressed in an ‘*moi dokei*’ or a ‘*phainetai*’ (‘it seems so’, ‘looks like it’),⁹⁸ it does not follow that to ignore this difference is sophistry. Indeed, we might now have an approach to Socrates’ practice that not only extenuates him for his disregard of this difference, but suggests that the disregard is in fact *essential* to the practice and its potential benefits to its participants.

In a serious game of chess, as we noted, the fact that you moved your bishop *reluctantly* gives me no reason at all not to take your exposed queen: indeed, not to do so would be patronising or hubristic, or a sign that I was playing with less than complete seriousness, or at any rate, that I felt something less than respect for my opponent. Sometimes one might play this way with a child who is learning how to play, but what is distinctive of that context is precisely the fact that the learner is

Beverluis, *Cross-Examining Socrates*,” *Oxford Studies in Ancient Philosophy* 20 (2001): 303 rightly notes, Beverluis here ‘articulates an unease ... which others have felt but not spelt out so clearly.’

⁹⁸ Again, the point is put succinctly in Beverluis, *Cross-examining Socrates*, 45-6: ‘the phenomenon of giving (or withholding) one’s assent is not as straightforward as it seems. Assent is not an either/or affair. There are *degrees* of assent, ranging from the unqualified ... to the more circumspect ... to the still less accommodating ... to the downright grudging Except for “Certainly” and its semantic equivalents, all these locutions signal *qualified* assent; and qualified assent is not assent *simpliciter*. ... Socrates’ interlocutors do not typically “assert categorically” the premises proposed for their assent; their responses are usually more qualified and guarded.’ Beverluis reaches a damning indictment of Socratic practice; far from revealing inconsistent beliefs in his interlocutors in the course of his putatively ‘peirastic’ procedures, he merely shows that there are tensions between propositions that only a tendentious reformulation allows him to attribute to them (see especially 57–8).

not yet ready to take full responsibility for her moves. As she gets better, there will be less and less call for her teacher to go on doing this. Indeed, some teachers may never play at anything below their best level; this has all the advantages, and some of the dangers, of a swimming instructor whose method involves pushing a non-swimmer into the pool at the deep end.

Another feature of the chess game is the fact that every move involves a decision. Occasionally in a game, it will be clear that only one move is really possible and there will be little need for extended deliberation before the decision. More often, every move will involve extensive reflection on a host of probabilities. Still the move itself is what it is. In the best case, it will reflect the player's judgment about what move was most rational; often – certainly in amateur games – the decision will be somewhat arbitrary, at least between two or three *prima facie* attractive possibilities between which one is unable to distinguish in the time available. One is liable to get it wrong, but a decision must be made; the clock is ticking.

The Socratic *elenchus* is a similar sort of game, a – what else to call it? – *language* game. It is by design a cooperative game, but only succeeds in being one in practice if both players want it to be. As in chess, there are two sides, and a set of permissible moves – roughly, those permitted by the usual norms of consistency and rationality.

When it is one's turn to make one's move, one must come to some decision, ideally one based on rational considerations, on what move one is prepared to make, and that can mean, what moves one is prepared to live with the consequences of having made. But these consequences are themselves uncertain; one cannot know in advance what they will be. One only knows that one is playing with a very canny player, and that every weakness in one's moves will soon be exposed.

Unlike in chess, the player of the Socratic language game has not an opponent but a partner. If he has an opponent that he must resist, it is reality itself and its stubborn refusal to let itself be known easily, and this is something the players confront together.⁹⁹ The Socratic *elenchus* is especially unlike chess – perhaps in this it is more like a chess problem than a game of chess – in allowing a player to withdraw a move once it is shown to lead to difficulties,¹⁰⁰ and try another move instead, in light of the weaknesses of one's first move, often more in hope than in confidence that this time one will get it right. More can be said about this analogy – about whether a chess game can test moves or only the player's skills, whether the *elenchus* can help one achieve not only consistency but truth. These are old

⁹⁹ A remark of Socrates' to Critias in the *Charmides* strongly suggests such a picture: 'don't you believe it to be for the common good, or for that of most men, that the state of each existing thing should become clear?' (*Charmides*166d)

¹⁰⁰ A clear example of Socrates permitting emendations is *Republic* I 340b. Clitophon makes a suggestion to Thrasymachus about how he might avoid certain of Socrates' objections. Polemarchus protests: 'That isn't what he said', but Socrates is nonchalant: 'It makes no difference, Polemarchus ... If Thrasymachus wants to put it that way now, let's accept it.'

questions about the *elenchus* and the analogy will not by itself help us solve them, though it does help us to put the old questions in new, and potentially more fruitful, ways.

One such question is this: why does Socrates treat his interlocutors' sometimes obviously reluctant assent as criterial of their beliefs? Given the apparatus developed in the last few sections, we can now offer this simple answer: he does not. The things they are inclined to assert or assent to – enthusiastically, hesitatingly, or grudgingly – reveal the state of their soul, not just the linguistically articulated states that deserve the name of belief or judgment. But it is only when the internal divisions of their soul are forced into a form sufficiently determinate and precise for them to be legitimate moves in this game that the game can proceed.

In other words, I am suggesting that Socrates' interlocutors are not always best described as *having inconsistent beliefs*; this widespread designation of them is more an artefact of modern scholarship than a reflection of anything explicit in Plato. I suggest that we had better say that there are in their minds dispositions that cannot all be articulated so as to yield a set of consistent beliefs. This is not necessarily a problem for the interlocutor; one can lead a good life without rationalising all one's dispositions. But Socrates is not concerned with all their dispositions, only with

those pertinent to the ethical life, hence his focus on the dispositions of discourse and behaviour that constitute their grasp on the value terms that structure their ethical lives: courage, temperance, justice and so forth. There are constant hints in Plato's characterisation of Socrates' interlocutors that the messiness of their dispositions is costing them something – Charmides wakes up everyday with a headache (*Charmides* 155b), Euthyphro has alienated his own family and is the laughing-stock of his fellow citizens whenever he makes his heterodox religious pronouncements in the assembly (*Euthyphro* 3bc), Meno is too drawn to the overwrought register of a Gorgias to recognise a plain truth when he sees one (*Meno* 76c).

A further point may be made about Socrates' good faith. The *elenctic* exchanges often allude explicitly to their protreptic functions, in particular, the exhortation to continue to inquire, bravely and tirelessly, even when Socrates is not there to help.¹⁰¹ Some of his interlocutors do heed this advice, most do not, and some of Plato's earliest readers would have read his narratives of young men in

¹⁰¹ E.g. the closing words of the *Theaetetus* 210bc ('if ever in future you should attempt to conceive or should succeed in conceiving other theories, they will be better ones as the result of this inquiry'), and Socrates' remarks to Meno (*Meno* 81d) after he has set out the 'theory of recollection' as a response to his paradox – that one cannot inquire into what one doesn't know ('... nothing prevents a man ... discovering everything else for himself, if he is brave and does not tire of the search We must, therefore, not believe that debater's argument, for it would make us idle, and fainthearted men like to hear it, whereas my argument makes them energetic and keen on the search.') On the latter passage, see the helpful discussion in Dominic Scott, *Plato's Meno* (Cambridge University Press, 2006), 82 and *passim* on the *Meno's* protreptic elements.

conversation with Socrates knowing well exactly what become of those young men since.¹⁰² Socrates attempts to induct them into a practice that has the potential to improve them. We are now in a position to explain why the *elenchus* has this potential. That potential lies in its replicating in a discursive, speculative connection the conditions of responsibility and urgency more natural to practical deliberation. There is little in ordinary life to force us to put any work into forming opinions that can withstand rational scrutiny. A surprise examination might do it, or an imminent political referendum. Socrates is perhaps suggesting that we had better prepare ourselves for these surprises. Many of the young men with whom he converses will have an important part to play in the life of their city. The public stage is not a place in which to improvise an opinion or pick one arbitrarily;¹⁰³ for this reason among others, they would do well to pay some attention while they can to the state of their opinions – or rather, with the haphazard, vacillating thoughts they have been contented with, not even making the effort to turn them into opinions in the first place.

¹⁰² Debra Nails, *The People of Plato: A Prosopography of Plato and Other Socratics* (Indianapolis: Hackett Publishing, 2003) is full of poignant biographies of Socrates' interlocutors and their future careers.

¹⁰³ There is an oft-quoted, if unverified, story about Lord Mountbatten choosing a date for the transfer of power to an independent Indian government off the top of his head in order not to let a journalist get the impression he was not entirely in control of the process. If it is true, there could be – because of the horrific violence that was to accompany the Partition of India – no better example of the dangers of improvising opinions on significant matters.

Like a trainee lawyer at a moot court, or a cadet conducting a mock raid, Socrates' interlocutor is being given the chance to cultivate the virtues of character and intellect that will stand them in good stead in setting where there is the risk of provoking more than an expression of Socratic irony. These are serious exercises, and people are evaluated in them by standards not far from those they will encounter in the non-ersatz versions of these exercises. The simulacra of real responsibility for one's opinions in Socrates' *elenctic* exchanges might be seen as preparations for true responsibility, when one cannot assert things in jest and one has to live with the consequences of the things one says and does. Life outside the Socratic encounter, conversations with people other than Socrates, are in many ways less demanding. But outside the *elenctic* setting, one does not have the luxury – as it were – of saying things simply because one thinks them to be true; a host of pragmatic considerations are apt to obtrude. Socrates tries to create a space within which playfulness and experiment can be mingled with seriousness and responsibility. In a word, the *elenchus* is a game that is good for the soul as other games are good for the body. Hence the almost militaristic tone of Socratic protreptic – 'be bold', 'be confident', 'do not hesitate'. If I am right, these enjoinders are not things he says in addition to asking them to say what they believe; if I am right, they are the same enjoinder.

The analogy of soul and body is well attested in Plato, and in Greek culture more generally. But it has significant implications for how we understand Socratic practice. The opening scene of the *Charmides* draws attention to this analogy: when Socrates is informed of the beauties of Charmides' body, he replies that he should prefer to undress his *soul* and inspect that (*Charmides* 154e). In both cases, the inspectee – as it were – must show a certain willingness to be examined. In the one case, it involves a willingness to strip down and display one's body; in the other, it involves a willingness to participate in dialogue with Socrates.¹⁰⁴ The analogy will be shown to be even closer if one contrasts the physical inspection of a living person with a post-mortem autopsy: the inspection of the living body is not the inspection of any old physical object; it is the inspection of a physical object of a special sort, characterised by its animation, and therefore, by its activity. Stripping down in Plato is prelude to a wrestling bout. Even a standard doctor's examination will require some degree of co-operation from the patient – 'Say ah', 'breathe in' – beyond simply making one's body available. The state of a body, no less than that of the soul, is manifested in its activity. If Socratic practice has in it any place for

¹⁰⁴ See also *Protagoras* 352ab, *Theaetetus* 162ab and 169ab, and *Symposium* 216e6–7 for the recurrence of the stripping analogy.

Cartesian first-person privilege, it is not the privileged access implied by the perceptive model of self-knowledge; it is, rather, the privilege of an agent.¹⁰⁵

It is widely recognised that one of the benefits of the *elenchus* is the prospect of rising from the self-contradictions into which Socrates has forced one and strive for consistency in one's assertions. The usual way of putting this involves a reference to replacing false beliefs with true ones; I have been arguing that we would do better to speak of speaking of replacing internequine dispositions – to coin a phrase – with more harmonious, and therefore *stabler*, ones. The idea is that stability is a ideal inherent in what it is for something to be a belief – its semantics, its relation to the world, its epistemology. Part of the reason some content is classified as a belief rather than some other sub-doxastic state (as we should say) has to do with the extent to which it is stable.¹⁰⁶ This way of thinking about belief has its Platonic precedents, most notably in the passage of the *Meno* that compares beliefs, even true ones, to the fabled statues of the sculptor Daedalus:

... they ... run away and escape if one does not tie them down but remain in place if tied down. ... To acquire an untied work of Daedalus is not worth much, like acquiring a runaway slave, for it does not remain, but it is worth much if tied down, for his works are very

¹⁰⁵ In this, again, I am entirely of one mind with Woolf, “Socratic Authority”, 23: ‘the parallelism between soul and body as objects of examination should not be viewed simply as the innocent by-product of a pre-Cartesian way of thinking. It suggests, rather, a theoretical stance that provides vital underpinning for the Socratic mission.’

¹⁰⁶ A recent paper informed by the contemporary epistemological literature on sub-doxastic states that discusses Plato's views on belief is Jessica Moss, “Plato's Appearance-Assent Account of Belief,” *Proceedings of the Aristotelian Society* 114, no. 2pt2 (July 1, 2014): 213–38.

beautiful. What I am thinking of when I say this? True opinions. For true opinions, as long as they remain are a fine thing and all they do is good, but they are not willing to remain long, and they escape from a man's mind, so that they are not worth much until one ties them down by (giving) an account of the reason why. (*Meno* 97e-98a; cf *Euthyphro* 11b-e)

The context here is a contrast between true beliefs and knowledge, but there is a point here about the nature of beliefs in themselves; the inclination to assert them in virtue of which they count as one's beliefs in the first place is a fragile thing, apt to change in the face of unprecedented contexts where one is no longer inclined to assert them.¹⁰⁷ A stable belief is, among other things, one that persists over time. For an ethical project that uses the examination of beliefs as a way of examining a whole life, it is easy to see why stability in one's beliefs would also be part of an ethical ideal; life is, after all, a long-term project. This suggests that volunteering oneself repeatedly into discursive contexts that threaten to destabilise one's beliefs is important precisely as a device by which our dispositions of assertion and assent can be reified and made (relatively) consistent, to the point that they may justly be called our beliefs rather than mere propositional 'moods'.

This chimes well with, among other passages, the stretch of the *Meno* where the slave is interrogated about a simple problem in geometry. The aim of the passage in the dialogue is, on the face of it at least, an attempt to provide an argument for the 'doctrine of recollection', the idea that even at birth we possess all sorts of

¹⁰⁷ A helpful brief discussion of instability in belief that discusses the *Meno* passage is Irwin, *Plato's Ethics*, 143-5.

knowledge that the process of inquiry is really an attempt to remember. The exchange has Socrates comment at various points to Meno about the slave's progress, progress that is meant – whether or not Meno perceives it – to mirror Meno's own.

You realize, Meno, what point he has reached in his recollection. At first he did not know ...; even now he does not yet know, but then he thought he knew, and answered confidently as if he did know, and he did not think himself at a loss, but now he does think himself at a loss, and as he does not know, neither does he think he knows. ... we have probably achieved something relevant to finding out how matters stand, for now, as he does not know, he would be glad to find out ... (*Meno* 84ab)

The slave's initial confidence, like Meno's, was unfounded. But it is only by having him assert and assent unequivocally to a wide range of propositions that this could be shown to him – as a conductor might encourage choristers to sing loudly in rehearsals, the better to pick up their mistakes. Thus,

What do you think, Meno? Has he, in his answers, expressed any opinion that was not his own?

No, they were all his own.

And yet, as we said a short time ago, he did not know? – That is true.

So these opinions were in him, were they not? – Yes.

So the man who does not know has within himself true opinions about the things that he does not know? – So it appears.

These opinions have now just been stirred up like a dream, but if he were repeatedly asked about these same things in different ways, you know that in the end his knowledge about these things would be as accurate as anyone's. – It is likely. (*Meno* 85b–d)¹⁰⁸

¹⁰⁸ I have modified Grube's translation slightly.

I am not here concerned with the doctrine of recollection itself; Socrates' argument for it in the *Meno* is notoriously weak, the putative proof falling well short of demonstrating its truth. But this passage gives us the materials we need for the sort of account I have been arguing for. The slave's opinions have 'been stirred up like a dream'; long-standing dispositions of the slave's, dispositions that, whether or not they go back to the beginning of time, certainly date from earlier than his encounter with Socrates, have been articulated into simple statements to which he has, for the most part correctly, given his assent. Now the real work begins, though the slave, being a slave, was likely never given the chance to do so: to ask, and to be interrogated, again and again, in different ways, about this material, making connections between isolated propositions until they reveal themselves in a systematic, Euclidean, whole.

The solitary internal back-and-forth described in the *Theaetetus* – with its partial resemblance to a genuine inter-personal exchange – might be less than enough to effect such a cognitive transformation.¹⁰⁹ Solitary reflection is too open to devolve into self-deception or wishful thinking; the presence of an actual, rather than

¹⁰⁹ A different set of considerations on the analogies and disanalogies between internal and interpersonal deliberation is explored in A. G. Long, *Conversation and Self-Sufficiency in Plato* (Oxford: Oxford University Press, 2013), 109–38 and *passim*, see also Mary Margaret McCabe, "Form and the Platonic Dialogues," in *A Companion to Plato*, ed. Hugh Benson (Oxford: Blackwell Publishing, 2006), 51 ff.

merely notional, interlocutor allows reflection on what one believes – or indeed on what to *believe*¹¹⁰ – to be made to answer to public, and therefore systematically enforceable, norms, as opposed to the self-indulgent and dangerously accommodating way in which one enforces a merely private norm on oneself. The Socratic game is best for the soul when it is played with someone else.

7. Scorekeeping in a Socratic language game

For all that Socrates mostly converses with his interlocutors one at a time, it has long been recognised that his conversations serve the broader purpose of social criticism. In the typical case, they serve this critical purpose because the sorts of things his interlocutors are disposed to say are not unique to them. Even those who pronounce themselves heterodox (e.g. Euthyphro, Protagoras, even Callicles) can hardly be heterodox about everything. Their opinions, attitudes and habits, in a word their dispositions, are ones they have in common, to a considerable degree, with their contemporaries. Consequently, for Socrates to show a singular interlocutor to be mistaken is – when what is shown to be mistaken is the articulation of a *common* disposition – is for him to reveal a problem in (what we might well call) common-sense morality itself, at least insofar as that interlocutor

¹¹⁰ Note Socrates' phrasing: 'we have probably achieved something relevant to finding out how matters stand' (*Meno* 84b).

has done his best by it. That last qualification is essential. What Socrates manages to show of an interlocutor is not (or rarely) that he is wrong, *simpliciter*, but that his assertions – and therefore his acquired commitments – are inconsistent. But this need not be an indictment of common-sense morality if there is reason to think the interlocutor failed to be good spokesman for it. There is always a further question to ask about a Socratic refutation: is there really no better argument for the claim putatively refuted than the one given here? This question suggests another: if Socratic interlocutors, however conventional, can only ever speak for themselves, can an exposé of their inconsistencies ever amount, by itself, to social criticism?

There are one or two places where this question is raised in the dialogues themselves. One of them is the stretch of the *Protagoras* we have briefly discussed earlier where Socrates first imposes, then relaxes, the requirement that his interlocutor say only what he believes. Protagoras finds it difficult to offer simple assent to Socrates' question about justice and piety being things of the same sort:

“It’s not so absolutely clear a case to me, Socrates, as to make me grant that justice is pious, and piety just. It seems a distinction is in order here. But what’s the difference? If you want, we’ll let justice be pious and piety just.”

“Don’t do that to me! It’s not this ‘if you want’ or ‘if you agree’ business I want to test. It’s you and me I want to put on the line, and I think the argument will be tested best if we take the ‘if’ out.” (*Protagoras* 331c)

But a short while later, Socrates seems to take a different attitude. He gets Protagoras to recognise that he is required (presumably by the demands of

consistency) to make a choice between abandoning either the claim that “for one thing there is only one opposite” or that “wisdom is different from temperance”

(*Protagoras* 333a). “So,” he asks,

... does someone who acts unjustly seem temperate to you in that he acts unjustly?”

“I would be ashamed to say that is so, Socrates, although many people do say it.”¹¹¹

“Then shall I address myself to them or to you?”

“If you like, why don’t you debate the majority position first?”

“It makes no difference to me, provided you give the answers, whether it is your own opinion or not. I am primarily interested in testing the argument, although it may happen both that the questions, myself, and my respondent wind up being tested.” (*Protagoras* 333bc; cf *Charmides* 155c–e)

Passages such as these seem to suffer from a tension between an inquiry conceived in peirastic, and therefore deliberately *ad hominem*, terms and one aimed at an examination of claims independently of their provenance. The difference between these two conceptions is captured in the fact that in the first sort of inquiry, the acceptability of a claim to an interlocutor is sufficient to justify its being treated as a premise in further argument; but the second would call for a premise to be evaluated by some other independent criteria of plausibility, e.g. whether some argument can be given for it. This tension can be resolved if we note the important fact that in both cases Socrates continues to require answers of his interlocutor,

¹¹¹ I am assuming the usual translation of ‘*epei*’ here as ‘although’, rather than – as Irwin, “Say What You Believe”, 8–9 argues – ‘since’.

though in the *Protagoras* passage at least, they need no longer be answers they themselves endorse. Why the continued reliance on the interlocutor's answers?

There is indeed one interlocutor who simply cannot see the point of this. Calicles in the *Gorgias*, now disinclined to participate in the discussion, asks the obvious question: 'Couldn't you go through the discussion by yourself, either by speaking in your own person or by answering your own questions?' Socrates, with reservations, agrees:

Let's by all means do it that way then. I suppose that all of us ought to be ... eager to know what's true and what's false about the things we're talking about. That it should become clear is a good common to all. I'll go through the discussion, then, and say how I think it is, and if any of you thinks that what I agree to with myself isn't so, you must object and refute me. For the things I say I certainly don't say with any knowledge at all; no, I'm searching together with you so that if my opponent clearly has a point, I'll be the first to concede it. (*Gorgias* 505d-506a; again, see also *Charmides* 155c-e)

Several familiar elements occur in this exchange – the conception of the discussion as a joint inquiry, and the continued requirement that any disagreement be registered rather than passed over in silence. If the respondent has been dispensed with, it is not because Socrates wants the chance to speak without interruption.

We might entertain the following explanation of why Socrates continues to require a respondent, however uncommitted: Socrates seems to want to show that the conclusions he derives follow from premises that are not merely idiosyncratic

opinions of his own; if they be derivable from what his interlocutors are willing to own as their beliefs, all the better. If not, then perhaps they can be shown to be derivable from a set of propositions attributable to some less specific entity, such as what Socrates and Protagoras call the position of the ‘majority’ or ‘many’, what we have been calling common-sense morality. But it will still need a spokesperson purporting, or at least trying, to speak on its behalf; how else can their position enter into the dialectic?¹¹²

Common-sense morality exists, if it exists anywhere, in the minds – or as Socrates would have it, souls – of the members of the community whose common sense it

¹¹² I take it to be evident that it would have a valuable contribution to make. Irwin, “Say what you believe”, 11-2, rejects Vlastos’s suggestion (Gregory Vlastos, “The Socratic Elenchus,” *Oxford Studies in Ancient Philosophy* I (1983), 38) that Socrates is trying to spare Protagoras the embarrassment of owning to such a position, and argues instead, I think rightly, that Socrates welcomes the opportunity to discuss a view that would not have been aired if the requirement that Protagoras only say what he believes were rigorously enforced. On my interpretation, this must be restated as follows: Socrates sees that Protagoras’ disinclination to own to sharing the majority position will prevent it having the hearing it certainly deserves; so it would be best for Protagoras to speak for that position without taking on any personal commitments. Jacques A. Bailly, “What You Say, What You Believe, and What You Mean,” *Ancient Philosophy* 19, no. Special (1999), 74 helpfully puts the interpretative crux thus: ‘Plato asserts the import of the independence of an argument or position from its originator as well as the dependence of the position on its originator’; sometimes, as we have seen, he does both within short stretches of text. Bailly’s intriguing explanation involves appealing to Plato’s authorial intentions and his affectation of sincerity in the dialogues. My explanation is more economical: the truth or falsity of a (non-indexical) proposition is independent of the identity of its originator; on the other hand, the costs and benefits of believing (or coming to believe) some proposition can hardly be made sense of without some account of that proposition’s compatibility with other propositions one might be inclined to endorse. As such, one needs a respondent as a focus for assessing these costs and benefits.

is.¹¹³ It will manifest itself in the discursive and behavioural dispositions of those members, and might in that form prove amenable to articulation, analysis, and assessment. But it is no easy task speaking for common-sense morality. For one thing, common-sense morality may not in fact be unitary or coherent. It may have multiple strands which even on the most charitable interpretation will not be resolved into a wholly consistent set of propositional articulations.¹¹⁴ Further, one's limited experience of life or society might leave one unable to say with any authority where common-sense morality stands on some question. Those with philosophical inclinations, often isolated or alienated from the majority, might be particularly prone to distorting its attitudes.¹¹⁵ These are familiar ideas, and continue to be aired in the context of contemporary discussions of the status and authority of 'intuitions' in philosophical argument.

¹¹³ A version of this point, though he does not use the phrase 'common-sense morality', is well made by Frede, "Dialogue Form", 215: 'Beliefs about these subject-matters, like virtue, reality, justice, evil, do not form relatively small, isolated clusters; they form sheer endless chains which, and this is of equal importance, determine, or help to determine, our whole life and the life of the society we live in. We are brought up with them, they help to form our character and our general outlook and attitude. They help to determine where we see our interests, they shape our ambitions.'

¹¹⁴ Some of these contending strains in the common-sense morality of Plato's time are described in Terence Irwin, "Common Sense and Socratic Method," in *Method in Ancient Philosophy*, ed. Jyl Gentzler (Oxford: Oxford University Press, 2001), 47. See also Irwin's discussion of the problems this fact generates for anyone claiming, as Plato's Socrates sometimes seems to do, to be offering a coherent interpretation and development of ideas from common-sense morality (*passim*).

¹¹⁵ This point is put wryly in Kenneth Dover, *Greek Popular Morality in the Time of Plato and Aristotle* (Indianapolis: Hackett Publishing, 1974), 7: 'In the light of modern philosophers' assertions about what "we" say, think and feel, I am inclined to treat [Plato and Aristotle's statements about what "most people" say and think] with caution.'

Of course, no individual member of the community can be assumed to be a wholly reliable spokesperson, for the reasons canvassed above. But it is only if common-sense morality is filtered through the dispositions of an individual respondent to Socratic questioning that it can become the subject of the familiar sort of examination. Some reasons for this are implicit in my claims in earlier sections: if common-sense morality is to get a fair hearing, someone needs to take ownership of it, stand behind the dispositions that constitute it. The ‘many’ do not need to be present – how could they be? – for the examination. If their spokesman shows evidence of inconsistency, it suggests (though it does not prove) that common-sense morality itself is to that degree inconsistent. Reliable reportage is hard enough, as we have seen, when the system of dispositions one is asked to articulate are one’s own. In being asked to speak for the majority, Protagoras has – if his attitudes are as heterodox as he sometimes claims – taken on an even more formidably *imaginative* project, to speak with authority for the dispositions of others.

Protagoras cannot manage it for long, and the conversation moves on to other things. A much more systematic attempt to speak for ‘the many’ appears in the *Republic*, and it is to its first two books that I will turn in the remainder of this essay. The long conversation reported in the *Republic* is motivated by the questions nicely

summarised in the remarks of Socrates' that conclude Book I: What is justice? Is it a virtue or a vice? Is it profitable for its possessor? These questions are not posed baldly right at the beginning of the dialogue (as they are in the *Meno*, for instance). The *Republic*, rather, creeps up on its motivating questions, questions that emerge more or less organically from a conversation that does not start out with any obvious philosophical promise. The dialogue opens at the Piraeus, Athens' port town, where Socrates has come to observe the festival of the Thracian goddess Bendis. The young men who congregate around him take him to the house of one of their number, Polemarchus, and Socrates enters into a conversation with Polemarchus' aged father Cephalus. Their conversation, not usually the subject of extensive commentary because of the more obvious philosophical riches of the exchange with Thrasymachus to follow, is rich with pathos and irony.¹¹⁶

Cephalus appears from a sacrifice he has been performing in the courtyard 'sitting on a sort of cushioned chair with a wreath on his head' (*Republic* 328bc), an image of conventional piety. He greets Socrates warmly: 'you ought to come here more often, for you should know that as the physical pleasures wither away, my desire for

¹¹⁶ The most extensive recent scholarly treatment of the exchange, attentive to both historical, literary and philosophical nuance, is Mark Gifford, "Dramatic Dialectic in *Republic* Book 1," in *Oxford Studies in Ancient Philosophy*, ed. David Sedley, vol. XX (Clarendon Press, 2001), 25–106.

conversation and its pleasures grows' (*Republic* 328d). Socrates replies with apparent warmth:

I enjoy talking with the very old, for we should ask them, as we might ask those who have travelled a road that we too will probably have to follow, what kind of road it is, whether road and difficult or smooth and easy. And I'd gladly find out (*puthoimēn*) from you what you think about this, as you have reached the point in life the poets call "the threshold of old age" (*epi gēraos oudōi*). Is it a difficult time? What is your report about it (*exangelleis*)? (*Republic* 328de)

Socrates seems to be sincerely interested in what Cephalus might have to tell him about aging. On this subject at least, the old man is in an authoritative position (consider the epistemic implications of the verb 'exangelleis', suggesting the 'exangeloi' of Greek tragedy who brings news of off-stage events). Cephalus for his part is pleased to be asked:

By god, Socrates, I'll tell you exactly what I think. A number of us, who are more or less the same age, often get together in accordance with the old saying. When we meet, *the majority* (*hoi ... pleistoi*) complain about the lost pleasures they remember from their youth, those of sex, drinking parties, feasts [...], and they get angry as if they had been deprived of important things and had lived well then but are now hardly living at all. *Some others* moan about the abuse heaped on old people by their relatives, and because of this *they repeat over and over* that old age is the cause of many evils. But I don't think they blame *the real cause*, Socrates, for if old age were really the cause, I should have suffered in the same way and so should everyone else of my age. (329ab, emphases added)

There are many fascinating things in this opening exchange. Take the two formal flourishes: Socrates' self-conscious use of the poeticism 'the threshold of old age' (the earliest surviving appearances of the phrase are in the *Iliad*)¹¹⁷ and Cephalus' counter-Homerism ('the old saying' to which he alludes is likely 'God ever draws

¹¹⁷ *Iliad* XXII 60, XXIV 487.

together like to like', which appears in the *Odyssey*).¹¹⁸ These are quotations unmoored from their (likely) sources. Cephalus' opening remarks move from the formulaic invocation of a proverbial platitude to reporting on the contrasting elements in the experience of his contemporaries. What is the cause of the evils characteristic of old age? Is it the absence of sensual pleasures? Is it the abuse of relatives? Or is it, as Cephalus claims to think, 'the way people live'? He continues:

But as it is, I've met some who don't feel like that in the least. Indeed, I was once present when someone asked the poet Sophocles; "How are you as far as sex goes, Sophocles? Can you still make love with a woman?" "Quiet, man," the poet replied, "I am very glad to have escaped from all that, like a slave who has escaped from a savage and tyrannical master." I thought at the time that he was right, and I still do, for old age brings peace and freedom from all such things. When the appetites relax and cease to importune us, everything Sophocles said comes to pass, and we escape from many mad masters. In these matters and in those concerning relatives, the real cause isn't old age, Socrates, but the way people live. If they were moderate and contented, old age, too, is only moderately onerous; if they aren't, both old age and youth are hard to bear. (*Republic* 329b-d)

We get in these lines yet another quotation, this time from a contemporary, Sophocles. The habitual conversational device of conversing with proverb, quotation and anecdote, common among educated Athenians, will be familiar to students of the surviving texts from that period. Plato's characters quote constantly.¹¹⁹ The claim to representational verisimilitude apart, the point of the quotations, for Plato's readers, is to mark the characters in the dialogues as belonging to a certain class within a certain society and culture. Within the dramatic context of the dialogues themselves, the quotations are revealing both of

¹¹⁸ '...ὄς αἰεὶ τὸν ἁμοῖον ἀγεὶ θεὸς ἡὸς τὸν ἁμοῖον' (*Odyssey* xvii. 218).

¹¹⁹ For a general discussion of the practice, see Dorothy Tarrant, "Plato's Use of Quotations and Other Illustrative Material," *The Classical Quarterly* 1, no. 1-2 (January 1951): 59-67.

the personality of the quoter and of the conventions of intelligent conversation that Socrates is often trying to subvert. Most importantly of all, they give expression to what all the parties to the conversation treat as part of a shared corpus of quasi-authoritative utterances, worthy at the minimum of discussion. The characters in the *Republic*, like the characters in all Plato's dialogues, are deeply embedded in a culture, unlike Hylas, Philonous, Cleanthes, Philo and other deracinated philosophers of the early modern dialogue tradition. Plato's characters are steeped, or have at least waded repeatedly, in their culture's canonical texts. Their conversations with each other are, and cannot help being, continuations of a wider cultural conversation that gives any particular conversation its background, significance, the terms with which a line of thought is to be pursued, and a point of departure for the aspiring dissenter.

This is most obvious in the opening line of the *Meno*: Meno's question ('Can virtue be taught?') did not originate with him. It was a question very much part of the ethical discourse of the wider Greek society, the sort of subject on which the Athenian analogue to the man on the Clapham omnibus was as likely as not to have an opinion: compare our 'Is there a God?', 'Creation or evolution?', 'Nature or nurture?' The *Meno* is an attempt to answer an old question. So is the *Republic*, but – as we shall see – its questions do not come already articulated as Meno's was. The first

book and a half of the *Republic* are attempts to bring a half-understood intellectual unease to the surface and give it clear expression. The route to its foundational questions about justice and its value involves a slow trek that starts in a conversation among old people in response to a different, and less foundational, question: What explains the characteristic pains of ageing?

Cephalus has given his answer: it is the way people live, moderately or immoderately, that determines what one's old age will be like. Socrates responds:

When you say things like that, Cephalus, I suppose that *the majority of people don't agree*, they think that you bear old age more easily not because of the way you live but because you're wealthy, for the wealthy, they say, have many consolations.

That's true; they don't agree. And *there is something in what they say*, though not as much as they think. (*Republic* 329e, emphases added)

The first italicised phrase casts Cephalus as a familiar type in Plato's dialogues: the self-conscious holder of a minority view, a view at variance with common sense.

The second italicised phrase makes an important concession: the majority view has something to commend it. It is true, Cephalus allows, that a 'good person wouldn't easily bear old age if he were poor, but a bad one wouldn't be at peace with himself even if he were wealthy' (*Republic* 330a). Cephalus comes close to saying that wealth might be a necessary condition for a tolerable old age; it certainly doesn't make it harder. But he denies the claim he finds implicit in the majority view, that wealth is sufficient. Further, he has a different idea of why wealth can be such a consolation

in old age. Old age, he says, is an intimation of mortality, and therefore, of the imminence of punishment in the Hades of the well-known stories. As such, it is a spur to reflection about one's conduct. The wealthy – if they are also 'decent and orderly' (*Republic* 331a) – are likelier never to need to 'cheat or deceive someone against our will and from having to depart for that other place in fear because we owe sacrifice to a god or money to a person' (*Republic* 331ab).

Socrates' response to these remarks is, in dramatic terms, a turning point. From these discursive remarks of Cephalus', he manages to educe what looks like a definition of justice, something Cephalus never purported to give:

A fine sentiment, Cephalus, but, speaking of this very thing itself, namely, justice, are we to say unconditionally that it is speaking the truth and paying whatever debts one has incurred? Or is doing these things sometimes just, sometimes unjust? I mean this sort of thing, for example: Everyone would surely agree that if a sane man lends weapons to a friend and then asks for them back when he is out of his mind, the friend shouldn't return them, and wouldn't be acting justly if he did. Nor should anyone be willing to tell the whole truth to someone who is out of his mind. (*Republic* 331bc)

The manoeuvre would be, in an ordinary conversation, at least odd, if it were not so familiarly Socratic. At no point in his remarks did Cephalus ever purport to be offering a definition of justice. Why does Socrates turn an exchange with an old man waxing eloquent about old age into an opportunity to launch an inquiry into the nature of justice?

I propose that we understand Socrates' remark here (*Republic* 331bc) – in the language of modern pragmatics – as his *registering the conversational score*.¹²⁰ Even if Cephalus never affirmed a definition of justice, his remarks were *presupposing* some understanding of justice. They were certainly affirming a great many things explicitly: Cephalus' allusion to telling the truth and paying one's debts appeared as part of his answer to Socrates' quite specific question: 'What's the greatest good you've received from being very wealthy?' (*Republic* 330d). His answer, boiled down to its essence, appears to be this: the greatest good I have received from being wealthy is freedom from fear of punishment in the afterlife for my injustices. This claim is only intelligible in the light of his claim that 'Wealth can do a lot to save us from having to cheat or deceive someone against our will and from having to depart for that other place in fear because we owe sacrifice to a god or money to a person' (*Republic* 331b).

Cephalus is hedging, claiming that 'wealth *can...*' be helpful in these ways, not that it always is – but this is understandable enough; he was only asked what goods *he* received from being wealthy. But by his own lights, the value of wealth for him lies in the fact that possessing it obviates, or at any rate reduces, the need for him to

¹²⁰ The idea that conversation – or 'discursive practice' – may helpfully be modelled along the line of a baseball game, with a sort of mental scoreboard keeping track of explicit statements, presuppositions, common ground, etc dates back to the classic paper by David Lewis, "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8, no. 1 (January 1979).

commit injustice. This is only a tenable claim if his wealth reduced (we might say) his 'net' need to commit injustice: it is possible, indeed plausible, that having great wealth might increase the need to commit injustice. But this could not be the case if injustice were *defined*, and not merely characterised partially, in terms of deception under financial duress and debt. If there were nothing more to injustice than these things, then there might be something to Cephalus' claim that the value of money to him might be understood in terms of its ability to ward off these specific ills. But his deep concern is with the consequences of injustice. It is, then, crucial to the viability of Cephalus' point that his characterisation of injustice be more or less exhaustive. If there turns out to be more to injustice than deception and unpaid debts, and conversely more to justice than truthful business-dealing and solvency, then Cephalus' answer to Socrates' question is a piece of self-deception. It is certainly true that Cephalus had asserted no definition; but he needs his claims about justice and injustice to be roughly exhaustive. Hence, Socrates' question: 'are we to say unconditionally that [justice] is speaking the truth...' To take him to have committed himself to a definition of justice is then a demand of interpretative *charity*, not a rude interposition of Socratic priorities into a conversation that has no space for them.

It might be urged that Socrates' education of a definition of justice from Cephalus' meandering remarks violates some other principle of conversational propriety. The point may be made in general terms, as a claim about when it is proper to urge considerations of rational cogency against an interlocutor's utterances, or as a specific claim about Socrates' shift from the deference of his first remarks to Cephalus ('I'd gladly find out from you what you think about this', *Republic* 328e) to the more critical questioning of *Republic* 331bc. To put the charge crudely, having lulled Cephalus into a false sense of security by affecting to take his thoughts on ageing as authoritative, Socrates unexpectedly, and unfairly, springs a new set of dialogical constraints on him.

This is unfair to Socrates. Socrates' first, polite, question to Cephalus made it clear that he sought a first-personal, narrative, account from him of the experience of ageing. Cephalus, however, answered him in a quite different style. Far from recounting stories, he instead took the conversation straightaway into the stuff of debate, explanation and abstraction. He mentioned rival views purporting to account for the pains of ageing, and committed himself explicitly to a view that explained those pains in terms of 'the way people live' (*Republic* 329d). It is not often noticed that when Socrates, later on in *Republic* I, utters those famous words, 'it is

no ordinary matter we are discussing, but how we are to live' (352d), he is referring back to a phrase from Cephalus.

We might then say that Socrates' initial remarks introduced two presuppositions: one about the subject-matter of the conversation, the other about the sort of conversation it was to be. Socrates appeared willing, perhaps contrary to his own predilections, to have a conversation premised on Cephalus' superior epistemic authority about ageing. It was Cephalus' remark about 'the way people live' that made the conversation one about ethical matters, and his exposition of his view in contradistinction to those of his peers, that made it one about *explaining* (rather than simply describing) the pains of ageing. The conversational score was steadily shifting, largely as a result of Cephalus' willing contributions to it, towards the presuppositions characteristic of the traditional Socratic dialogue: an exchange structured by reason, argument, and the search for definitions. It should not have come as a surprise to anyone when Socrates took the natural next step.

Of course, Cephalus does not stay in the conversation long; he has his sacrifices to attend to.¹²¹ But his son Polemarchus is willing to stay and defend – take ownership

¹²¹ Stanley Cavell's suggestive remark on Cephalus (*Must We Mean What We Say?: A Book of Essays* (New York: Cambridge University Press, 2002), xxviii) is worth quoting here: 'Those who guard themselves

and responsibility for – the claim.¹²² He insists that the definition his father was unwilling to defend is right, ‘if indeed we’re to trust Simonides at all’ (*Republic* 331d). A couple of variations on the earlier definition are tried out, but the familiar *elenctic* exchange is disrupted when Thrasymachus, a rhetorician who has been listening to the conversation with rising impatience, bursts out and pronounces the whole affair ridiculous, demanding that Socrates say what he thinks justice is. Socrates demurs and asks Thrasymachus to say what justice is, since he evidently has something to say. His answers (justice is the good of the stronger, justice is following the law, justice is the good of another, and so forth), which seem less concerned with definition than with debunking,¹²³ are united by an emphasis on the ways in which justice involves the sacrifice of one’s own interests and is, as such, very much a second best. Halfway through their conversation, there is this intriguing exchange:

Do you really include injustice with virtue and wisdom, and justice with their opposites?

I certainly do.

That’s harder, and *it isn’t easy now to know what to say* (*touto ... ēdē stereōteron ... kai ouketi rhaidion echein hoti tis eipē*). If you had declared that injustice is more profitable, but agreed that it is a vice or shameful, as some others do, we could have discussed the matter on the

from philosophy show a healthier respect for it than those who are certain they know its results and know to whom they apply.’

¹²² This in fact gives us a nice parallel with the early passages of the *Theaetetus*, where the aged Theodorus hands over the conversation to his willing student. The key word here is *willing*: Cephalus and Theodorus are both, no doubt for different reasons, unwilling to risk the sort of exposure and vulnerability that comes of conversing with Socrates. Willingness, here as elsewhere in Plato, seems to be a trait associated with youth.

¹²³ For a defence of this point, see Rachel Barney, “Socrates’ Refutation of Thrasymachus,” in *The Blackwell Guide to Plato’s Republic*, ed. Gerasimos Xenophon Santas (Oxford: Blackwell Publishing, 2006), 44–62; also Williams, “Plato Against the Immoralist” in *The Sense of the Past*, 98.

basis of conventional beliefs (*kata ta nomizomena legontes*). But now obviously (*dēlos*), you'll say that injustice is fine and strong and apply to it all the attributes we used to apply to justice, since you dare to include it with virtue and wisdom.

You've *divined my views* exactly (*alēthestata ... manteuēi*).

Nonetheless, we mustn't shrink from pursuing the argument and looking into this, just as long as I take you to be saying what you really think (*heōs an se hupolambanō legein haper dianoōi*). And I believe that you aren't joking now, Thrasymachus, but are saying what you believe to be the truth (*ta dokounta peri tēs alētheias legein*).

What difference does it make to you, whether *I* believe it or not (*eite moi dokei eite mē*)? It's *my account* (*logon*) you're supposed to be refuting.

It makes no difference. But try to answer this further question... (*Republic* 348e–349a)

Socrates seems to have got the measure of Thrasymachus' position, to the point that he can guess what else he is likely to say. Thrasymachus' advocacy no longer matters as far as the details of his position are concerned. Socrates proceeds to conduct his usual procedures, eliciting Thrasymachus' assent to a range of propositions, and then revealing a surprising consequence of those assertions:

Then, a just person has turned out to be good and clever, and an unjust one ignorant and bad.

Thrasymachus agreed to all this, not easily as I'm telling it, but reluctantly, with toil, trouble, and—since it was summer—a quantity of sweat that was a wonder to behold. And then I saw something I'd never seen before—Thrasymachus blushing. But, in any case, after we'd agreed that justice is virtue and wisdom and that injustice is vice and ignorance, I said: All right, let's take that as established. But we also said that injustice is powerful, or don't you remember that, Thrasymachus?

I remember, but I'm not satisfied with what you're now saying. I could make a speech about it, but, if I did, I know that you'd accuse me of engaging in oratory. So either allow me to speak, or, if you want to ask questions, go ahead, and I'll say, "All right," and nod yes and no, as one does to old wives' tales.

Don't do that, contrary to your own opinion.

I'll answer you so as to please you, since you won't let me make a speech. What else do you want?

Nothing, by god. But if that's what you're going to do, go ahead and do it. I'll ask my questions.

Ask ahead. (*Republic* 350c–e)

This is a puzzling moment. Whatever does Socrates think he is doing here when he says 'if that's what you're going to do, go ahead and do it'? It is amply evident that he is singularly failing – not that he can be really trying by this point – to *persuade* Thrasymachus. But he is offering arguments, and valid ones. Are they sound? Thrasymachus can hardly think so; at any rate, his assent does not imply that he does. Whether they are sound or not will depend on what there is to be said for their premises, and the premise that does the most work in Socrates' arguments in Book I is the one that claims for justice the status of a virtue that brings harmony to the soul that possesses it and therefore enables it to perform its function well. He does not argue this point at any length in Book I, if at all; hence the apologia of one interpreter that Socrates' 'arguments work against Thrasymachus, despite their obvious faults ... because those faults betray the overcompression of deep truths'.¹²⁴ For Socrates to show that the arguments are sound, persuasive to Thrasymachus or not, he will have to make good this claim on his behalf by offering some argument for his central premises.

¹²⁴ Nickolas Pappas, *Routledge Philosophy Guidebook to Plato and the Republic* (Abingdon: Routledge, 2002), 50.

As it happens, Socrates will be forced to do just this. Book I ends with the image of Thrasymachus having some claim to a moral victory – Socrates has ‘won’, but only because Thrasymachus had stopped playing the game. But Book II begins with the surprising admission: ‘I thought I had done with the discussion, but it turned out to have been only a prelude. Glaucon showed his characteristic courage (*andreiotatos* – etymologically, ‘manliness’) on this occasion too and refused to accept Thrasymachus’ abandonment of the argument’ (*Republic* 357a).

There are goods valued for their own sake, he says, such as enjoyment; there are other goods valued for what comes from them, such as physical exercise; and there are the best goods of all, those that valued *both* for their own sake and because of what come from them – paradigmatically, seeing, knowing, and being healthy (paraphrasing *Republic* 357a–d). In which of these classes of goods does justice belong? Socrates would have it classified among ‘the finest goods’, with seeing, knowing and being healthy.¹²⁵ Glaucon replies:

That isn’t how it seems to most (*ou toinun dokei ... tois pollois*); [justice seems to them to belong] rather to the onerous kind, and is to be practised for the sake of the rewards and popularity that come from a reputation for justice, but is to be avoided because of itself as a burden. (*Republic* 358a)

¹²⁵ The best recent discussion of Glaucon’s classification is Andrew Payne, “The Division of Goods and Praising Justice for Itself in *Republic* II,” *Phronesis* 56, no. 1 (January 1, 2011): 58–78.

The mention here of ‘how it seems to most’ evokes three familiar Platonic contrasts: appearance and reality, opinion and knowledge, the many and the wise. Much of the *Republic* will be devoted to an explication of these dichotomies and spirited arguments asserting their right-hand side, but this is one of their early appearances in the text. Adeimantus and Glaucon’s challenge to Socrates – itself an attempt to improve on Thrasymachus’ challenge in Book I – poses the question that Socrates spends the rest of the *Republic* answering. But Glaucon and Adeimantus pose their challenge in a spirit distinct from that in which Thrasymachus posed his. They keep their challenge at an arm’s length, cast themselves self-consciously as *advocati diaboli*: ‘It isn’t ... that I believe any of that myself,’ says Glaucon. ‘I’m perplexed, indeed, and my ears are deafened listening to Thrasymachus and countless others’ (*Republic* 358c); some among these countless others are named in the speeches the brothers are shortly to make.

The brothers speak at some length, quoting from a variety of sources, about why they think justice, for all that has been said on the subject, still wants a defence that grounds its value in something other than the value of the rewards that conventionally accrue to a person with a reputation for it. They don’t want to believe what they are shortly to say about the value of justice, but in the absence of a good argument for the contrary position, they might find themselves falling into

it. The perplexity to which Glaucon confesses is the motivating force of the *Republic*. He and Adeimantus want to uphold a position to which they are drawn, and that has the support of a portion of the popular discourse on the subject. But that discourse has another strand. Few will quite come out champions of the Thrasymachean position; but the conventional injunctions recommending justice over injustice are, by this point in Athenian history, coming to acquire the flavour of sententious platitudes, with little in the way of argument to commend them to a reflective Athenian. The more a reflective person attends to these platitudes and tries to derive something like a *theory* of justice from them, the likelier he is to find those platitudes failing to yield a coherent set of claims about the nature of justice. Most alarming of all, the platitudes when considered together are compatible with a recommendation of a convincing simulacrum of justice as they are with a recommendation of justice itself.

Not for the last time in history, we have an ‘uneasy gap ... between the spirit of the theory itself and the spirit it supposedly justifies’.¹²⁶ The spirit these remarks seek to commend is one of disinterested love of virtue; the commendations themselves, however, are not disinterested. It is this uneasy gap that seems to inspire Glaucon and Adeimantus’ perplexity, a gap symptomatic of a problem in common-sense

¹²⁶ The remark, whose original context is a critical discussion of Sidgwick’s ‘indirect utilitarianism’, is from Bernard Williams, *Ethics and the Limits of Philosophy* (Abingdon: Routledge, 2006 [1985]), 108.

morality. The gap between the content of its platitudes and the terms in which those platitudes are argued for is evidence that the entity in question is not entirely stable when it is the subjection of sustained reflection. What the brothers' subsequent remarks and most of the Platonic corpus will reveal is something more serious: a body of opinions and attitudes underlying a society's customary practices riven with internal tensions of which members of that society are coming to be conscious. Faced with an unstable common sense, there is a strong temptation to adopt the kind of cynicism about ethical discourse and practice to which Thrasymachus, among others, gives dangerously eloquent voice. What makes this instability a *problem* – in the ordinary, rather than philosophical, sense of the word – is the fact that even such a pair as Glaucon and Adeimantus, well-off and well-versed in the canonical texts of their culture, are more readily able to offer an argument, and a sophisticated one at that, for the cynical position they affect, no doubt sincerely, to scorn, than they are able to adduce considerations in favour of the position they are actually inclined to hold. This is not the sign of an ethical culture in good health.¹²⁷

When Glaucon and Adeimantus are finished presenting their case against justice, Socrates finds himself sharing Glaucon's perplexity, as we saw earlier:

¹²⁷ My argument here is, as will be noted, in some measure indebted to Alasdair MacIntyre, *After Virtue: A Study in Moral Theory* (University of Notre Dame Press, 1981), 131–45.

you must indeed be affected by the divine if you're not convinced that injustice is better than justice and yet can speak on its behalf as you have done. And I believe that you really are unconvinced by your own words. I infer this from the way you live, for if I had only your words to go on, I wouldn't trust you. (*Republic* 368ab)

The brothers do not live as if they believe what they have been saying. Their conduct is, as yet, undisturbed by their sceptical doubts. But for how long will their faith in the superiority of true justice to mere dissembling survive the onset of doubt? Their practice wants a theory to vindicate it, and their society's common sense – as embodied in the canonical utterances the brothers quote as part of their case – does not give them one. They look to Socrates for reassurance.¹²⁸ If this characterisation is along the right lines, the *Republic* may be seen an enactment of, and a response to, a tension in the common-sense morality of the era of Athenian history in which it is set. That tension needs spokesmen to articulate it, and when Glaucon and Adeimantus step up, the words seem to pour out of them. It is almost as if they have been waiting a long time to voice their doubts, and who else is there to voice them to but Socrates? They claim to have suspended belief about justice itself, but this leaves a vacuum in their souls that – if traditional ideas are simply inadequate – Thrasymachus stands ready to fill. (What, after all, is Thrasymachus

¹²⁸ These passages of the *Republic* were no doubt on Bernard Williams's mind when he wrote, in a passage that refers explicitly to the *Gorgias*, that the justification for morality sought by a philosopher 'is in fact designed for the people who are largely within the ethical world, and the aim of the discourse is not to deal with someone who probably will not listen to it, but to reassure, strengthen, and give insight to those who will' (Bernard Williams, *Ethics and the Limits of Philosophy* (Abingdon: Routledge, 2006 [1985]), 26).

doing at the house of Cephalus if not looking for new students? It would certainly explain his mounting rage during Socrates' early exchange with Polemarchus.)¹²⁹

Socrates continues:

The more I trust you, however, the more I'm at a loss (*aporō*) as to what to do. I don't see how I can be of help. Indeed, I believe I'm incapable of it. And here's my evidence. I thought what I said to Thrasymachus showed that justice is better than injustice, but you won't accept it from me. On the other hand, I don't see how I can refuse my help, for I fear that it may even be impious to have breath in one's body and the ability to speak (*eti empneonta kai dunamenon phthengesthai*) and yet to stand idly by (*apagoreuein*) and not defend justice when it is being prosecuted. So the best course of action is to give justice any assistance I can. (*Republic* 368bc)

The tables have been turned, and, as so rarely happens, Socrates needs to do what he is always exhorting his interlocutors to do: to be bold, to stand behind his position and do his best to defend it. He introduces a new metaphor, but it has little to do with midwifery or recollection:

... I told them what I had in mind: The investigation we're undertaking is not an easy one but requires keen eyesight. Therefore, since we aren't clever people, we should adopt the method of investigation that we'd use if, lacking keen eyesight, we were told to read small letters from a distance and then noticed that the same letters existed elsewhere in a larger size and on a larger surface. We'd consider it a godsend, I think, to be allowed to read the larger ones first and then to examine the smaller ones, if they really are the same. (*Republic* 368cd)¹³⁰

¹²⁹ I am grateful to Myles Burnyeat for this observation.

¹³⁰ I have edited the Grube/Reeve translation slightly to reflect the fact that the last phrase 'if they really are the same' is a presupposition of the analogy, not the point of making it. Grube/Reeve's 'to see whether they really are the same' seems to me to be philosophically untenable: after all, the small letters in this analogy are too small to be read.

The metaphor is one of seeing, and of learning to see more clearly, and it inaugurates the analogy of city and soul that occupy the rest of the *Republic* and its investigation of individual and political justice and injustice. In the eight-and-half books to follow, not once does Socrates need to ask his interlocutors to say what they believe. For them and for Socrates, the question is now explicitly, *what are we to believe?* To answer this, they will learn little by looking ‘into themselves’. Their eyes must be, ‘so to speak, ... directed outward – upon the world’.¹³¹

8. Concluding remarks

A brief word on my choice of epigraph. Foucault’s remark (‘There is something in critique which is akin to virtue’) was made in a lecture linking the projects of twentieth-century critical theorists with that of Kant in his account of the nature of Enlightenment. But the reference to virtue teasingly suggests that critique might have an even earlier genealogy, reaching back to Socrates, and part of my aim in this essay has been to assemble the materials to show some ways in which this might be the case. I began by asking why Socrates sometimes insists that his interlocutors say only what they believe. I have argued the injunction is not a call to introspect, but rather to reflect, in good faith, about the facts of the matter.

Sometimes, their present reflections and the dispositions they bring into the

¹³¹ Evans, *Varieties of Reference*, 225.

conversation will point in the same direction; sometimes, they will not. When they do not, it takes a sort of courage to own to the fact. Socrates asks them what view they are genuinely willing to stand behind; he asks them, in other words, to adopt a *critical* stance. The courage it takes to do this is, I submit, the sense in which critique is akin to virtue: it is the ability to place the self in a position where it is able to take in how things might really stand, including how things stand with that self itself.

Viewed in this light, the *Republic* is, like Descartes's *Meditations*, or Kant's *Critique of Pure Reason*, a critical text in the deepest sense. It begins in common sense, with everyday practice, but does not assume that common sense is a fixed, static entity, readily available to the consciousness. Rather, it recognises that common sense is always in the making, always an incomplete project. Glimpses of the basic structure of common sense can be had, but not by introspection; common sense needs, instead, a context that makes it salient and allows the individual in whom it inheres to give it expression in speech and action. The *Republic* then invites its readers – as Socrates invites Glaucon and Adeimantus – to step back from that practice and consider what might need to be true of reality in order for that practice not to be exposed as ridiculous in the light of sceptical reason. Step by step, it attempts to reconstitute that practice, or at any rate some of it, in the light of what it finds to be

the truth about reality. Athens will not have looked the same when Socrates and the brothers walked home from the Piraeus.

Part II

‘An acute and wary sense of the ordinary’: Common Sense and Critique in the Philosophy of Bernard Williams

We are growing year by year more introspective and self-conscious: the current philosophy leads us to a close, patient, and impartial observation and analysis of our mental processes We are growing ... more unreserved and unveiled in our expression: in conversations, in journals and books, we more and more say and write what we actually do think and feel, and not what we intend to think or should desire to feel. We are growing also more sceptical in the proper sense of the word: we suspend our judgment much more than our predecessors, and much more contentedly: we see that there are many sides to many questions: the opinions that we do hold we hold if not more loosely, at least more at arm's length: we can imagine how they appear to others, and can conceive ourselves not holding them.

Henry Sidgwick¹³²

But it can not be granted that a man may make a judgment about man. Existentialism spares him from any such judgment. The existentialist will never consider man as an end because he is always in the making.

Jean-Paul Sartre¹³³

¹³² Henry Sidgwick, *Miscellaneous Essays and Addresses*, ed. Eleanor Mildred Sidgwick and Arthur Sidgwick (London: Macmillan, 1904), 60.

¹³³ Jean-Paul Sartre, *Essays in Existentialism*, ed. Wade Baskin (New Jersey: Citadel Press, 1965), 61.

1. Descriptive and Revisionary Ethics

The preface to Peter Strawson's *Individuals: An Essay in Descriptive Metaphysics* (1959) introduces the now well-known contrast between 'descriptive' and 'revisionary' metaphysics with a deceptively simple formulation: 'Descriptive metaphysics is content to describe the actual structure of our thought about the world'; it aims 'to lay bare the most general features of our conceptual structure'. By contrast, 'revisionary metaphysics is concerned to produce a better structure.'¹³⁴ Strawson's conclusion makes a connection between our conceptual structure and our metaphysical beliefs, held at whatever level of sophistication: 'It is difficult to see how such beliefs could be argued for except by showing their consonance with the conceptual scheme which we operate, by showing how they reflect the structures of that scheme.'¹³⁵

These programmatic remarks raise questions that are best answered not by any further elaboration, but by offering some examples of this kind of metaphysics, and of this kind of metaphysical argument; indeed it is how Strawson tried to answer them in that book. But some basic questions remain about the project of a descriptive metaphysics: how widely does Strawson's first-person pronoun range? Who, in other words, are the 'we' whose conceptual scheme is the subject matter of

¹³⁴ P. F. Strawson, *Individuals: An Essay in Descriptive Metaphysics* (Methuen, 1959), 9.

¹³⁵ Strawson, *Individuals*, 247.

the inquiry? How determinate a thing is a conceptual scheme anyway? And what exactly it is for a belief to ‘reflect the structures’ of our conceptual scheme?

Some of these questions acquire a deeper resonance when they are posed about

Strawson’s distinction when it has appeared outside its original metaphysical home.

Derek Parfit writes in the preface to his *Reasons and Persons* (1984):

Strawson describes two kinds of philosophy, descriptive, and revisionary. Descriptive philosophy gives reasons for what we instinctively assume, and explains and justifies the unchanging central core in our beliefs about ourselves, and the world we inhabit. I have great respect for descriptive philosophy. But, by temperament, I am a revisionist. ... I try to challenge what we assume. Philosophers should not only interpret our beliefs; when they are false, they should change them.¹³⁶

Parfit’s paraphrase comes close to suggesting, by implicature and omission, that

Strawson denied that final claim; and it attributes to him a view about philosophy in

general, rather than just metaphysics. A great deal rests here on how we understand

‘the unchanging central core in our beliefs about ourselves’; in fact, it is not entirely

clear where Parfit’s own project fits into Strawson’s schema. As A. W. Moore has

asked,

Suppose we are inclined to say one thing and a metaphysician urges us to say the very opposite. Is that revisionary because the metaphysician is challenging what we think? Or is it non-revisionary because the metaphysician is acceding to the concepts we use? Is the metaphysician acceding to the concepts we use? Perhaps saying the very opposite of what we are currently inclined to say would be so revolutionary that, if any of us did that, he or she would have to be interpreted as using old words to express new concepts These are familiar philosophical quandaries.¹³⁷

¹³⁶ Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), x.

¹³⁷ A. W. Moore, *The Evolution of Modern Metaphysics: Making Sense of Things* (Cambridge: Cambridge University Press, 2011), 13.

They are indeed, and the question – descriptive or revisionary? – can profitably be pressed about any philosopher, not as an end in itself, but as a way of making better sense of what that philosopher is doing. Strawson’s schema can help to motivate a question apt to puzzle readers of one of the twentieth century’s most significant moral philosophers, Bernard Williams. In reading Williams, one is apt to be puzzled by the presence of two rival strands. A revisionary strand runs through those writings of Williams’s that excoriate the many philosophical confusions and theoretical excesses he attributes to the ‘morality system’.¹³⁸ It appears in his defences of a kind of political liberalism from its conservative critics,¹³⁹ in his commitment to the British Labour Party, and his public interventions in debates on such divisive questions as the regulation of pornography and drugs.¹⁴⁰ It is above all to be found in his championing of Nietzsche’s scepticism about much ordinary moral thought.¹⁴¹

There is, however, also a *descriptive* – and, perhaps in consequence, conservative – strand to be traced through his *oeuvre*. It is to be found in his consistent resistance

¹³⁸ The central exposition of this notion is Bernard Williams, *Ethics and the Limits of Philosophy* (Abingdon: Routledge, 2006 [1985]), 174–96.

¹³⁹ E.g. his reviews of such historians as Christopher Lasch and Maurice Cowling, reprinted in Williams, *Essays and Reviews*, 169–78.

¹⁴⁰ A. W. Moore, “Sir Bernard Arthur Owen Williams (1929 - 2003)”, in *Oxford Dictionary of National Biography*, 2007 offers a brief account of Williams’s public role.

¹⁴¹ For an early instance of such scepticism, see his remarks about Nietzsche in a review of Iris Murdoch’s *The Fire and the Sun*, reprinted in Williams, *Essays and Reviews*, 142–5.

to the radical overhauls of common-sense morality implied by the theoretical constructions of philosophers in the Benthamite and Kantian traditions, and the seriousness with which he takes such notions as integrity, shame and honour, notions which those constructions threatened to render obsolete or otiose.¹⁴² It will be found in his emphases on a philosophical practice more attentive to the phenomenology of ethical experience, its sentimental no less than its cognitive components.¹⁴³ Something of its trace will be detected in his occasional remarks on the allure of a kind of romanticism involves in much ordinary moral thought.¹⁴⁴ And it will be found in his defences of a kind of historically sensitive political ‘realism’ aimed at understanding and guiding political practice ‘now and around here’, as against an Anglo-American academic orthodoxy characterised by ‘liberal moralism’.¹⁴⁵

¹⁴² E.g. Bernard Williams, ‘A Critique of Utilitarianism’ in J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (Cambridge University Press, 1973), esp 97–100 and *passim*; Bernard Williams, *Truth and Truthfulness: An Essay in Genealogy* (Princeton: Princeton University Press, 2002), 115–7.

¹⁴³ Williams, *Ethics*, 93.

¹⁴⁴ E.g. Bernard Williams, *Morality: An Introduction to Ethics* (Penguin 1972), 92–5; ‘Practical Necessity’ in *Moral Luck* (Cambridge: Cambridge University Press, 1981), esp 130–1; *Shame and Necessity* (Berkeley: University of California Press, 1993), 75–77; Bernard Williams, *Moral Luck* (Cambridge: Cambridge University Press, 1981), 37–8; *Truth and Truthfulness*, 199–205.

¹⁴⁵ See ‘Realism and Moralism in Political Theory’ in Bernard Williams, *In the Beginning Was the Deed: Realism and Moralism in Political Argument* (Princeton: Princeton University Press, 2008), esp 8 and *passim*.

To put it bluntly: what was he doing, describing or revising? Was he with the champions of tradition, folk wisdom and common sense, or with their critics? Philosophers through history have tended to find themselves on one side or the other. If Williams's attitude was, as it seems, complicated, then it would help us to understand his philosophy better to have his commitments on this question clarified. Williams was primarily a moral philosopher even if he published extensively on questions in other branches of philosophy. The question about where he stood with respect to common sense is thus a question of where he stood with respect to a specific part of common sense, the part of it that has come to be labelled, at least since Sidgwick, 'common-sense *morality*'. Given just how much of the philosophical activity of his colleagues was taken up with the task of discovering the true moral theory which would offer an authoritative account of the grounds for judging an action (e.g.) right or wrong, it is easy to see why Williams might be thought to have harboured (suspect) conservative inclinations. If moral philosophy just was the quest for the true moral theory, then to be hostile to that task was to be hostile to moral philosophy itself. And if one was not some manner of nihilist – and Williams was not evidently a nihilist – then one was left defending common-sense morality – understood by its critics (and some defenders) as a body of inherited prejudices along with the wilful unreflectiveness necessary to sustain a commitment to them. Once put in this way, it is clear that this presents us with a

false choice: why need it be a choice between nihilism, unreflectiveness and theory?

Could there not be some way of thinking, reflectively and rigorously, about ethics that does not presuppose that the only, or best, way of being constructive in ethics is to offer a systematic theory? If there could be such a style of ethics, what attitude to common-sense morality would it involve?

Although he belonged to, and for the most part identified with,¹⁴⁶ a tradition of philosophical writing that prides itself on its clarity and transparency, Williams has been found both to need and to reward exegesis. He needs it because his style, while often elegant and precise, is condensed to a fault and passes with unwonted speed over the hardest passages of argument.¹⁴⁷ He rewards it because there almost invariably emerges from his suggestive but unelaborated remarks on any subject a substantial, provocative, and often attractive view, or more often a *question*, that he did not trouble to spell out in so many words.

For the most part in this essay, I am concerned with exposition rather than evaluation. Williams had a long and productive career, and much that he wrote,

¹⁴⁶ See his remarks on style in Williams, *Ethics*, vii–viii.

¹⁴⁷ As A. W. Moore notes in his explanatory notes to the most recent edition of *Ethics and the Limits of Philosophy*, ‘Williams never belabours the obvious; and he rarely makes explicit what he takes to be implicit in something he has already said. His writing is therefore extremely dense. It leaves an enormous amount of work for the reader’ (*Ethics*, 204).

especially for non-specialist audiences, has not been easily available; as a result, the predominant focus has been on a small and unrepresentative set of texts and themes. I shall not try to defend every claim of Williams's; I hope, rather, to make it clearer how they cohere with each other. If there is an architectonic here, it is this: I offer a reading of Williams's writings as united by their aspiration to a *critical* moral philosophy – I use the word 'critical' in something like its Kantian sense – of attending to the structure and limits of moral experience, and the conditions of the possibility and legitimacy of such experience. I begin by finding the motivations for such an idea of moral philosophy in Williams's intellectual formation among the British thinkers of the 1950s in whose midst he spent his early career. I then propose that Williams's ethics may be read as a kind of 'liberation' philosophy, aimed towards helping the reflective reader to confront and overcome elements of her experience of morality that do not survive philosophical reflection. I go on to offer a close reading of Williams's two celebrated mini-narratives about 'George' and 'Jim' and offer an unorthodox account of Williams's aspirations in writing them. I conclude with a brief discussion of his prose style and what it reveals about his project more generally.

2. Williams and his Peers

The conferral on a thinker of canonical status does not guard him from – indeed it makes him specially vulnerable to – the distinctive misinterpretations of his philosophy perpetrated by historians of philosophy working in a certain, ‘analytic’, style. So keen can such interpreters be to locate arguments in the text that can then be evaluated for soundness that some important facts can be, effectively, forgotten: namely, that all philosophers are to some degree the product of their milieu.¹⁴⁸

Even if one’s historical methods are primarily analytic – that is to say, oriented towards the reconstruction of arguments independently of their social context – it can help the reconstruction of arguments to be sensitive at least to the intellectual elements of those milieu. No philosopher argues in a vacuum; a good rational reconstruction needs a reconstruction of the precursory dialectic, and Williams, that most reactive of philosophers, benefits greatly from being placed in a dialectical context. There is no natural limit on how far this contextual reconstruction might be taken. Given that the self-conception of many modern western philosophers places them in a more-or-less continuous tradition reaching back to the ancient Greeks, any part of this tradition may legitimately be thought to

¹⁴⁸ A good summary of debates on method in the history of political thought is David Runciman, “History of Political Thought: The State of the Discipline,” *The British Journal of Politics & International Relations* 3, no. 1 (April 1, 2001): 84–104, who concludes that the basic claims of the so-called ‘Cambridge school’ of intellectual historians, insofar as they deem significant for understanding a philosophical text the political and socio-cultural contexts in which it was written, are no longer seriously challenged.

form part of their dialectical context. I shall have to set up some (ultimately arbitrary) limits, and shall restrict myself to five influential senior contemporaries of Williams': J. L. Austin, Gilbert Ryle, R. M. Hare, Iris Murdoch and Elizabeth (G. E. M.) Anscombe; I shall also refer in passing to the influence of earlier figures, such as Wittgenstein, Nietzsche and Sidgwick. I shall restrict myself to tracing his influences and targets and will not arbitrate his often complex disagreements with these figures. My primary aim here is to give an account of Williams as a figure whose concerns emerge in part from the concerns of his contemporaries in the early decades of post-war Britain; in particular, I have been anxious to quote from and discuss at some length occasional writings of his from this period that have long been neglected.¹⁴⁹

From 1947 to 1951, Williams was an undergraduate at Balliol College, Oxford, where he studied for Oxford's iconic 'Greats' degree, which combines the study of Greek and Roman literature and history with the study of philosophy, ranging from the pre-Socratics to the latest in contemporary philosophy (*modulo*, of course, the prejudices and predilections of one's tutors). Williams received a 'congratulatory first' – a rare examination honour – and was elected to a prestigious Prize Fellowship at All Souls College, which he took up in 1953 having done his two years

¹⁴⁹ It is to be hoped that the recent republication of these writings (Williams, *Essays and Reviews*) will draw attention to them.

of (then still mandatory) National Service as a fighter pilot in Canada. The university had by then fully recovered from the privations of the war and its best thinkers were back from war work. The major philosophical event of Williams's undergraduate years at Oxford was the publication in 1949 of Gilbert Ryle's *The Concept of Mind*. R. M. Hare's *The Language of Morals* was published in 1952, shortly after Williams had graduated. 1952 was also the year that J. L. Austin took up the White's Chair in Moral Philosophy. Williams was well acquainted with all three philosophers, in particular with Hare who had been his tutor in philosophy at Balliol. Hare, Ryle, and Austin had different temperaments, and their work dealt with different subjects, but they had enough in common to be thought to be at the centre of a distinctive Oxonian style of philosophy; at least the latter two, and Austin in particular, would come to be described as practitioners of something called 'ordinary language philosophy'. There was enough in both the practice and the occasional methodological pronouncement of these thinkers to warrant the label, and I see no reason to place it in inverted commas in the following pages.

The strand of ordinary language philosophy that concerns us here is its alleged conservatism about language use. Indeed, the label has often been used as a pejorative, and to saddle the putative members of this school with commitments

they had done little to merit.¹⁵⁰ Yet, there was enough in the writings of the philosophers in question to see what the accusers might have been thinking of. The first figure worth considering is the one whose influence on Williams – if only as *provocateur* – was strongest: R. M. Hare, his undergraduate tutor, and predecessor in the White's Chair at Oxford, which he took up in 1990. Much of Williams's career was devoted to the spirited critique of views in, and about, ethics that were held in their most concentrated form by Hare.¹⁵¹ One such view is to be found in Hare's unapologetic circumscription of his subject matter in his Preface to *The Language of*

Morals:

Ethics, as I conceive it, is the logical study of the language of morals. ... I have deliberately avoided references to the problems of moral psychology. ... This is not because I consider these problems unimportant, nor because I have nothing to say about them, but because they are rather problems of the language of the psychology of morals, than of the language of morals itself.¹⁵²

¹⁵⁰ The best-known polemic against ordinary language philosophy was Ernest Gellner, *Words and Things: An Examination Of, and an Attack On, Linguistic Philosophy* (Victor Gollancz, 1959), which came with a laudatory foreword from Bertrand Russell, who complained that ordinary language philosophy believed 'common speech is sacrosanct, and that it is impious to suppose it capable of improvement' (xii). An early, and forceful, response to Gellner's thesis came from Michael Dummett, 'Oxford Philosophy' in *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press, 1978), 431–6, though Dummett thought Austin's influence on Oxford philosophy baneful ('although he was himself unquestionably a clever man, I always thought that the effect of his work on others was largely harmful, and therefore regretted the nearly absolute domination that for a time he exercised over Oxford philosophy' (xii)). More recent assessments of the 'movement', such as it was, are Avner Baz, *When Words Are Called For* (Cambridge, Mass: Harvard University Press, 2012) and Sandra Laugier, *Why We Need Ordinary Language Philosophy* (Chicago: University of Chicago Press, 2013).

¹⁵¹ Alasdair MacIntyre's review of his 1981 collection of papers *Moral Luck* noted wryly that 'if there is a book which more than any other constitutes the thesis to Williams's antithesis it is one published only after these papers had appeared in their original versions, R. M. Hare's [*Moral Thinking: Its Levels, Method and Point* (Oxford: Oxford University Press, 1981)]. If Hare had not existed, Williams should perhaps have seriously considered inventing him' (Alasdair MacIntyre, "The Magic in the Pronoun 'My,'" *Ethics* 94, no. 1 (October 1983): 113–25).

¹⁵² R. M. Hare, *The Language of Morals* (Oxford: Oxford University Press, 1952), iii–iv.

At no point in his career does Williams seem to have accepted such a conception of ethics – either in terms of its relationship with the study of a part of language, or in its exclusion of questions in moral psychology. Hare was almost certainly one of the targets of Williams’s widely quoted and openly provocative prefatory remarks to his first book, *Morality: An Introduction to Ethics*, when he spoke of

a particular charge against contemporary moral philosophy of the ‘analytical’ or ‘linguistic’ style: that it is peculiarly empty and boring. ... The emptiness of past works [in moral philosophy] has been the emptiness of conventional moralizing Contemporary moral philosophy has found an original way of being boring, which is by not discussing moral issues at all. ... the desire to reduce revealed moral commitment to a minimum and to use moral arguments in the role of being uncontentiously illustrative leaves an impression that all the important issues are off the page, and that great caution and little imagination have been used in letting tiny corners of them appear.¹⁵³

Hare’s later work took on a slightly different character, and he was happy to engage in substantive moral argument, and also to qualify the extent to which a focus on moral language committed one to any kind of philosophical conservatism. Still, Hare never ceased to hold that a description of certain necessary features of moral language placed limits on the form of ‘basic’ moral thought; combined with relevant empirical information, moral language could even determine the content of substantive moral principles. Hare was happy to declare in later life that ‘Common moral opinions have in themselves no probative force whatever in moral philosophy ... What does have probative force (directly in theoretical ethics, and

¹⁵³ Williams, *Morality*, 9–10.

only indirectly, and with the help of other factors, in morals) is the ordinary use of the moral words'.¹⁵⁴

A *festschrift* for Hare elicited these remarks from Williams:

Hare has always wanted, and earnestly wanted, moral philosophy to have a practical effect, to make a difference. He does not think it an obstacle to this that on his view moral philosophy is, roughly, a branch of philosophical logic. On the contrary: moral philosophy can make a difference only because it has authority, and it can have authority only because of its neutral status as a logic or linguistic subject. ... his present theory gives a very special explanation of what this authority is ... [:] that this neutral subject can yield foundations.¹⁵⁵

This suggested a deeper diagnosis about Hare's motivations, and his anxieties. In

Williams's view, Hare's reason

for thinking that philosophy should not proceed by reflection ... on our moral 'intuitions', opinions, or experiences ... is that they are merely ours ... the special form taken by his linguistic enquiry reveals that it is itself conditioned by the search for foundations, and by the desire to get away from what is merely 'ours'.¹⁵⁶

Hare's project in ethics, on this description, bears some resemblance to Strawson's in metaphysics. The focus on (some) moral language was supposed to lead one on to an apprehension of a structure that, as Strawson put it, 'does not readily display

¹⁵⁴ R. M. Hare, *Essays on Philosophical Method* (University of California Press, 1972), 122.

¹⁵⁵ Bernard Williams, *Philosophy as a Humanistic Discipline* (Princeton: Princeton University Press, 2008), 83. The essay first appeared, titled 'The Structure of Hare's Theory', in Douglas Seanor and N. Fotion, eds., *Hare and Critics: Essays on Moral Thinking* (Oxford: Clarendon Press, 1988) with a reply from Hare.

¹⁵⁶ Williams, *Humanistic Discipline*, 85. For a helpful discussion of Williams's evolving views on the question of whether ethics needed foundations, see Paul Sagar, "Minding the Gap: Bernard Williams and David Hume on Living an Ethical Life," *Journal of Moral Philosophy* 11, no. 5 (September 24, 2014): 615–38.

itself on the surface of language, but lies submerged.’¹⁵⁷ Williams’s most extensive treatment of the question of how far a linguistic method could take one in moral philosophy was in a chapter of *Ethics and the Limits of Philosophy* (1985) titled ‘The Linguistic Turn’. He was more than happy to allow that ‘moral philosophy, like other parts of philosophy, is properly concerned with reflection on what we say. Indeed, at one level it might have done better than it has if it had been more concerned with what we say’.¹⁵⁸ His objection to the sort of attention to moral language exemplified by Hare was its restricting itself to a narrow range of ‘thin’ ethical concepts such as ‘good’, ‘right’ and ‘ought’, concepts that Hare took – question-beggingly¹⁵⁹ – to be contained in the ‘thick’ concepts such as ‘kind’, ‘brave’ and ‘honourable’ in which much ethical discourse is conducted. However, to redirect one’s attention to these concepts raised its own set of problems. For one thing, thick concepts – defined by their combining the descriptive and prescriptive – are notoriously diverse, and rooted in particular cultures or subcultures. Yet, it

¹⁵⁷ Strawson, *Individuals*, 10.

¹⁵⁸ Williams, *Ethics*, 127. See also his remarks in ‘Morality and the Emotions’, his inaugural lecture at Bedford College, London in 1965 (Bernard Williams, *Problems of the Self* (Cambridge: Cambridge University Press, 1976), 207–8): ‘the diversity of what can be called reflection on language’ is equalled by the diversity of what can be called “the language of morality”, and there was no basic reason why a generous approach to the linguistic endeavour should not have embraced those features of our speech about morality that reveal or suggest the parts played by the emotions... [Linguistic moral philosophy] has imposed on the linguistic enterprise a concentration on the most general features of moral language ...’

¹⁵⁹ Williams seems to have held that the restriction could only be justified if thick concepts were, as Hare believed, reducible to thin ones. For a classic discussion of this question, see John McDowell, “Non-Cognitivism and Rule-Following,” in *Wittgenstein: To Follow A Rule*, ed. S. Holtzman and Christopher M. Leich (Routledge, 1981), 141–62; a helpful recent discussion of the ‘irreducibility thesis’ is Debbie Roberts, “Shapelessness and the Thick,” *Ethics* 121, no. 3 (2011): 489–520.

seems possible to understand the ethical concepts of other cultures without coming to use them oneself. Further, there is a question about the histories of such concepts and what it says about a society that it conducts its ethical evaluations in terms of the ‘honourable’ and ‘treacherous’ rather than in terms of the ‘right’ and the ‘good’.¹⁶⁰ Once these facts have been made salient, the analogy between ethics and metaphysics no longer seems a close one.

To the extent that Williams’s objection to the linguistic turn in moral philosophy was directed at the narrowness of the examples it theorised from, his concerns dovetailed with those of J. L. Austin, who memorably pronounced a certain position

a typically scholastic view, attributable ... to an obsession with a few particular words, the uses of which are over-simplified, not really understood or carefully studied or correctly described; and ... to an obsession with a few (and nearly always the same) half-studied ‘facts’. ... The fact is ... that our ordinary words are much subtler in their uses, and mark many more distinctions, than philosophers have realized ...¹⁶¹

¹⁶⁰ The difference between such societies lay in the fact – though this point is not elaborated in *Ethics and the Limits of Philosophy* – of their differing experiences of modernity and acceptance of its characteristic forms of rationality. For a more extended discussion of this point, see Williams, *Essays and Reviews*, 314–5 and *Ethics*, 130–31.

¹⁶¹ J. L. Austin, *Sense and Sensibilia*, ed. G. J. Warnock, 1st edition (Oxford: Oxford University Press, 1962), 3. The only explicit reference to Austin in *The Language of Morals* is this: ‘we can speak of “good art”, but not of “right art”, and of “good batting”, but not of “right batting”’; and on the other hand, we can say “You didn’t play the right note”, but “good” could not be substituted. As the work of Professor J. L. Austin has taught all who have had the benefit of instruction in his methods, such peculiarities may be – though they are not always – indicative of underlying logical differences’ (Hare, *Language*, 152). This is an Austinian point indeed, but it does not by itself absolve him of potential charges of a focus on certain items of vocabulary to the exclusion of relevant others.

The quotation is from Austin's *Sense and Sensibilia*, a posthumous reconstruction of material from his lectures where he used elements of ordinary language as part of an argument against 'sense data' theorists such as A. J. Ayer. If Williams's point came from a sympathy for Austin's claim about the subtlety of ordinary words allegedly neglected by Hare, then his criticisms of Hare acquire a markedly conservative character. That Austin was a sort of – qualified – conservative is strongly suggested by any number of his (frequent) methodological *obiter dicta*, of which we might consider one example below. The context is a critique of his opponents' insensitivity to the many uses of the word 'real' in ordinary speech:

Philosophers often seem to think that they can just 'assign' any meaning whatever to any word; and so no doubt, in an absolutely trivial sense, they can (like Humpty-Dumpty). There are some expressions ... which only philosophers use, and in such cases they can, within reason, please themselves; but most words are *in fact* used in a particular way already, and this fact can't be just disregarded. ... [T]here is certainly no reason why, in general, things should be left exactly as we find them; we may wish to tidy the situation up a bit, revise the map here and there, draw the boundaries and distinctions rather differently. But still, it is advisable always to bear in mind ... that the distinctions embodied in our vast and, for the most part, relatively ancient stock of ordinary words are neither few nor always very obvious, and almost never just arbitrary; ... that in any case, before indulging in any tampering on our own account, we need to find out what it is that we have to deal with; and ... that tampering with words in what we take to be one little corner of the field is always *liable* to have unforeseen repercussions in the adjoining territory. Tampering, in fact, is not so easy as is often supposed, is not justified or needed so often as is often supposed, and is often thought to be necessary just because what we've got already has been misrepresented.¹⁶²

Austin's remarks suggest some degree of conservatism, in that they claim more for inherited (linguistic) practice than the rampaging reformer is willing to see in it. He

¹⁶² Austin, *Sense and Sensibilia*, 62–3.

takes care to qualify the claim: ‘ordinary language is not the last word ... [but] it is the *first* word’.¹⁶³ It was this last point, with all its implications for philosophical practice were it to be taken seriously as a statement of method, to which Williams devoted the bulk of his review of *Sense and Sensibilia*, first published in the *Oxford Magazine*.¹⁶⁴ Even while praising ‘the fine linguistic ear which was one of [Austin’s] outstanding gifts’, he thought it ‘remarkable how little Austin use[d] his methods of close linguistic observation to support [his] specifically philosophical points’; ‘the linguistic observations seem[ed] often to be pursued for their own sake’.¹⁶⁵ However,

The linguistic study would lead back eventually to the subjects which the philosophical theorists had over-rapidly attacked – it was to be, in a famous phrase, the ‘begin-all’. How was the ‘begin-all’ to be pursued? Austin’s answer to this seems to have been a Baconian one: that one patiently assembled distinctions in ordinary usage, and then elicited from them a pattern or theory ...¹⁶⁶

This, Williams continued, raised its own problems: how could there be a finite and determinate *number* of distinctions unless the assembler had some pre-existing idea of what purpose they were to be put to? Or, as he put it in another place, how could the ‘begin-all’ ever end unless there was a theory that explained the grounds for, and limited the endless proliferation of, those distinctions, as the theory of evolution by natural selection gave a point to the otherwise largely whimsical

¹⁶³ See his remarks in J. L. Austin, “A Plea for Excuses: The Presidential Address,” *Proceedings of the Aristotelian Society*, New Series, 57 (January 1, 1956), 11 ff.

¹⁶⁴ For some helpful remarks on the importance of the *Oxford Magazine* as a venue for serious discourse in mid-20th-century Britain, see Stefan Collini, *Absent Minds: Intellectuals in Britain* (Oxford: Oxford University Press, 2006), 340. The review would certainly have been read by all his colleagues.

¹⁶⁵ Reprinted in Williams, *Essays and Reviews*, 42.

¹⁶⁶ Williams, *Essays and Reviews*, 42.

taxonomies of life-forms.¹⁶⁷ There was a further and more serious problem, suggested by the passage from Austin quoted above. Williams called it ‘a curious view of Austin’s, what might be called his “Wisdom of the Ages” thesis’.¹⁶⁸ Apart from the obvious fact that the thesis seemed to rule out – though in doing this it was hardly alone at the time – any philosophical positions that ‘tried to show that men were *quite mistaken* about some fundamental features of the world’, there was its alleged conservatism:

Now the charge that Austin was a linguistic conservative ... has been rebutted by reminding us that the study of existing uses was indeed a ‘begin-all’ – after the investigation, then possibly reform. But this is still, in fact, conservatism. It is particularly so if conjoined with the Baconian thesis about the methods of investigation, for on that thesis, the investigation will be literally interminable. Even without that, it ... amounts to the proposal that no revolution or even reform can be mounted without a thorough sociological investigation of the *ancien régime*; and this is a proposal which in other contexts is rightly regarded as amounting to the proposal that there be no reform.¹⁶⁹

There were further arguments. Ordinary language was continually being sullied by the incursion of innovations from ‘the despised hasty theorists’ in (e.g.) the sciences, so no sharp distinction between ordinary and technical language could be sustained; Austin’s methods were potentially self-defeating, after all, ‘the old is only

¹⁶⁷ Bernard Williams, ‘The Spell of Linguistic Philosophy [Interview]’ in Bryan Magee, ed., *Talking Philosophy: Dialogues with Fifteen Leading Philosophers*, (Oxford: Oxford University Press, 2001), 110 – 124, *passim*. Also Williams, *Humanistic Discipline*, 160–1: ‘In the philosophy of language ... the point has established itself that an isolated distinction or analysis lacks both sense and point: Austin’s professed view, that one collects linguistic distinctions like types of beetles, can be seen to be absurd about linguistic distinctions, and not very clever about beetles.’

¹⁶⁸ Williams, *Essays and Reviews*, 43.

¹⁶⁹ Williams, *Essays and Reviews*, 43.

what used to be new¹⁷⁰; they needed a better account of the significance of the mere antiquity of a distinction; above all, Austin needed to engage more effectively with ‘some real historical material about the changes in language’.¹⁷¹ Williams reached a damning conclusion, softened only slightly:

I suspect that the Wisdom of the Ages thesis is in fact a *myth*, a fanciful picture of the past designed to justify certain activities in the present. Austin was fascinated by, and extraordinarily gifted in detecting, nuances of ordinary speech. He also had great philosophical imagination and acumen. He sought to bring the two together, by urging the study of ordinary use as the method of philosophy, or at least a method necessarily preliminary to philosophy. ... When he ... succeeded, it was ... in his extremely skilful deployment of fine linguistic distinctions as a weapon to encourage – or, more characteristically, to discourage – the holding of certain philosophical beliefs.¹⁷²

There is no wholesale rejection of the ordinary language approach in these remarks, and some admiration for the talents of its most influential practitioner. But it is those distinctive talents that are the subject of his admiration, not the method itself. Austin, he said, ‘had not discovered ... a method that could be handed over to mechanics to churn out results’.¹⁷³ There was still a place in Williams’s mind for some of the theoretical ambitions Austin seemed to spurn. (As we shall see later, Williams is not referring here to the sort of *moral* theory he attacks elsewhere.)

Indeed, those ambitions might well be indispensable to any justification for turning

¹⁷⁰ Williams, *Essays and Reviews*, 44.

¹⁷¹ Williams, *Essays and Reviews*, 44. Cf Williams, *Ethics*, 131: ‘[the linguistic approach] is at least potentially closer to some understanding of the social and historical dimensions of ethical thought than some other approaches, which see it entirely in terms of an autonomous and unchanging subject matter. To draw attention to our ethical language can at least hold out the prospect of our coming to think about it, and about the ethical life expressed in it, as social practices that can change. The linguistic turn could have helped us, even if it has not actually done so, to recognize that ethical understanding needs a dimension of social explanation.’

¹⁷² Williams, *Essays and Reviews*, 44.

¹⁷³ Williams, *Essays and Reviews*, 43.

to distinctions in ordinary language in the first place. There is very little in Williams's published work to suggest that he was ever in thrall to Austin's way of doing philosophy.¹⁷⁴ Points of ordinary language crop up every now and then in his work, but they are rarely made to bear any special weight.¹⁷⁵

In this, he was closer to Gilbert Ryle than to Austin. Williams had great affection for Ryle and came to know him and his work well in his years at Oxford.¹⁷⁶ In a *London Review of Books* review of a posthumously published volume of Ryle's essays, he made several insightful observations about Ryle's progress – at any rate his movement – from the interest in the ('Continental') phenomenological tradition he had shown in his youth to the 'linguistic' style exemplified by *The Concept of Mind*:

¹⁷⁴ We have the testimony of his first wife Shirley Williams on this point – that Williams, very briefly, 'got caught up in the prevailing linguistic questions [at Oxford]. I think ultimately he saw that the culture was clever but shallow. But at the start he was rather dazzled by it ...' (Stuart Jeffries, "The Quest for Truth," *The Guardian*, November 30, 2002) Some of this ambivalence is nicely captured in a remark in a late essay, 'The Liberalism of Fear', where the unexpected and gratuitous lyricism stands at odds with the disapproval expressed by the remark itself: '... J. L. Austin ... coined the memorable phrase "linguistic phenomenology" but did not live to distinguish in his work between a phenomenology guided by language, which is what a lot of philosophy is, and an activity of which the point is less clear: the phenomenology of the use of language itself, the tracing of the shudder of an exquisite distinction' (Williams, *In the Beginning*, 53).

¹⁷⁵ E.g. a passing remark on the difference between the constructions 'A has reason to φ ' and 'There is reason for A to φ ' in 'Internal and External Reasons' (Williams, *Moral Luck*, 101); a brief note on Aristotle's observations about the ordinary language expressions used to denote the master-slave relation (Williams, *Shame and Necessity*, 106–7); a brief reminiscence of Austin's remark about 'the oddness of the sentence *I have a pain in my waist' (Williams, *Shame and Necessity*, 175 n. 6); a note on the ordinary language use of the phrase 'an experience' and its refinement by metaphysicians into an explanatory principle (Williams, *Humanistic Discipline*, 31); a brief discussion of how he took the word 'lie' to be standardly understood outside of philosophical contexts (Williams, *Truth and Truthfulness*, 96–7).

¹⁷⁶ His reminiscences of the man are very warm: see Bernard Williams, "Ryle Remembered," *London Review of Books*, November 22, 1979, reprinted in Williams, *Essays and Reviews*, 152–6 (esp 152 and 156).

He [Ryle] dealt in uses of words, and his arguments rested heavily on considerations of what did and did not make sense. Since he wrote in English, it was a question of what did and did not make sense in English, but he always dissociated himself from the minute interest in fine points of usage which some of his colleagues displayed. He claimed that any sound argument of the kind that he used reached below the level of a particular natural language, and could be translated.¹⁷⁷

Ryle's most important book was a sustained attack on various aspects of a received 'Cartesian' way of thinking about the mind, of which mind-body dualism was only one element. The other elements against which Ryle trained his guns were – as Williams nicely summarised it – 'theories which represented the mental life as a hidden immaterial process duplicating or paralleling observable doings', something which tended, for all Ryle's protestations to the contrary, 'in a behaviourist direction ... suggest[ing] that there was no conscious inner life at all'.¹⁷⁸

One of the many remarkable things about *The Concept of Mind* is the language in which Ryle described his aims. His introduction begins with a bluff, self-consciously commonsensical, almost anti-intellectual description of his starting point: 'We possess already a wealth of information about minds, information which is neither derived from, nor upset by, the arguments of philosophers.'¹⁷⁹ The tone here is that of a writer who has thoroughly cast off any obvious 'Continental' stylistic influences; the tutelary spirits of such a sentence are the (archetypically – or

¹⁷⁷ Williams, *Essays and Reviews*, 153.

¹⁷⁸ Williams, *Essays and Reviews*, 153.

¹⁷⁹ Gilbert Ryle, *The Concept of Mind*, 60th anniversary edition (Abingdon: Routledge, 2000), lix.

stereotypically – English) G. E. Moore of ‘A Defence of Common Sense’¹⁸⁰, itself part of an English tradition some have traced back to John Locke.¹⁸¹ The evidence that ‘we’ possess this wealth of information is simply our ability to use mental concepts with confidence and authority. Like in Austin, there is no serious consideration in Ryle of the possibility that ‘we’ could be *radically* mistaken in our use of mental concepts. But Ryle promptly circumscribes the scope of his apparent conservatism:

Many people can talk sense with concepts but cannot talk sense about them; they know by practice how to operate with concepts, anyhow inside familiar fields, but they cannot state the logical regulations governing their use. They are like people who know their way about their own parish, but cannot construct or read a map of it, much less a map of the region or continent in which their parish lies.¹⁸²

This is the well-known picture of philosophy as conceptual cartography.¹⁸³ This was conjoined to a historical thesis:

¹⁸⁰ Originally published in 1926, reprinted in G. E. Moore, *Philosophical Papers* (London: George Allen & Unwin Publishers, 1959), 32–59.

¹⁸¹ Ryle reports an amusing exchange with Bertrand Russell: “‘I think Locke invented Common Sense.’ To which Russell rejoined ‘By God, Ryle, I believe you are right. No one ever had Common Sense before John Locke – and no one but Englishmen have ever had it since’ (Gilbert Ryle, *Collected Papers: Volume I* (Hutchinson, 1971), 147). This self-conception in English philosophy is captured in Williams’s own (ironic) telling of the origin myth of Ryle’s tradition, a droll paraphrase of a narrative he finds in G. J. Warnock’s *English Philosophy since 1900*: ‘At the start we are presented with the macabre picture of British Idealism, intellectually corrupt, fraudulent ... Coming from foreign places, its rule was never more than the tyranny of occupation; and the rise of the hero, Moore, to drive it out is an affirmation not only of the light against the dark but of the native against the exotic. In the ensuing struggles, Moore has his allies, not always reliable: Russell, brave but unsteady; Positivism, secretly in love with the metaphysical enemy. [There follows the] epic of Wittgenstein, ... [an] extravagant figure ... With Ryle, victorious peace is almost achieved; and the story ends with Common Sense again on the throne, and the citizens of Oxford, calm but not idle, earning the unambitious rewards of honest toil’ (Williams, *Essays and Reviews*, 5–6).

¹⁸² Ryle, *Concept of Mind*, lx.

¹⁸³ This picture is briefly and helpfully discussed in Julia Tanney’s introduction to the Routledge issue of *The Concept of Mind* (Ryle, *Concept of Mind*, lv–lvii) and in Julia Tanney, “Gilbert Ryle,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2009, 2009, esp Section 2 (‘Philosophy as Cartography’).

it is part of the thesis of this book that during the three centuries of the epoch of natural science the logical categories in terms of which the concepts of mental powers and operations have been co-ordinated have been wrongly selected. Descartes left as one of his main philosophical legacies a myth which continues to distort the continental geography of the subject.¹⁸⁴

Ryle took it as his purpose to ‘explode’ the myth – which was ‘not to deny the facts but to re-allocate them’.¹⁸⁵ If he often fell into a polemical register in doing this, he thought it worth pointing out that

the assumptions against which I exhibit most heat are assumptions of which I myself have been a victim. Primarily I am trying to get some disorders out of my own system. Only secondarily do I hope to help other theorists to recognise our malady and to benefit from my medicine.¹⁸⁶

The therapeutic language here puts him in the immediate lineage of the late Wittgenstein, as does the idea that one might be the ‘victim’ of an ‘assumption’¹⁸⁷, in this case, one encouraged by aspects of the ‘mythical’ Cartesian picture of the mind and its relationship with the body. The medical metaphors here are no aberrations; Ryle returns to them at the end of the book. He acknowledges the virtues of the Cartesian picture, at least in its relative ‘productivity’ compared to the mechanistic

¹⁸⁴ Ryle, *Concept of Mind*, lx.

¹⁸⁵ Ryle, *Concept of Mind*, lx.

¹⁸⁶ Ryle, *Concept of Mind*, lxi.

¹⁸⁷ A couple of the relevant passages are §115: ‘A picture held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably’, and §299: ‘Being unable – when we surrender ourselves to philosophical thought – to help saying such-and-such; being irresistibly inclined to say it – does not mean being forced into an assumption ...’ (Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1953)). In Williams’s view, ‘Ryle shared ... with Wittgenstein, and no doubt in part derived from him ... the important belief that the philosophy of mind had to get beyond both dualism and behaviourism’ (Williams, *Essays and Reviews*, 155).

‘Hobbes-Gassendi story of the mind’¹⁸⁸ according to which ‘thinking just consists in making certain complex noises and movements’.¹⁸⁹ There follows a somewhat strange analogy: the Cartesians are like a company of soldiers who, finding the only available fortress to be weak, decide to stand in ‘the most fort-like thing they can see, namely, the shadow of the decrepit fort’.¹⁹⁰ This is not much better, he says, than ‘doctrinal homeopathy’, but it has the (important, solitary) virtue of recognising the disorder.¹⁹¹

There are places where Williams seems keen on some variety of just such a therapeutic conception of philosophy, but they must be seen in the light of his (more-or-less lifelong) pronouncements against it. A very early essay published in 1957, ‘Metaphysical Arguments’,¹⁹² briefly discusses a view of philosophy that he finds at least hinted at in Wittgenstein and elaborated in the work of John Wisdom, according to which – to simplify Wisdom’s views somewhat – the theories and arguments of metaphysicians were neurotic symptoms. The point was not to argue with them but to cure them by using ‘analytic [i.e. philosophical] technique to get to

¹⁸⁸ Ryle, *Concept of Mind*, 302.

¹⁸⁹ Ryle, *Concept of Mind*, 300.

¹⁹⁰ Ryle, *Concept of Mind*, 302.

¹⁹¹ Ryle, *Concept of Mind*, 303.

¹⁹² David Pears, ed., *The Nature of Metaphysics* (London: Macmillan, 1957); reprinted in Williams, *Humanistic Discipline*, 22–33.

the roots of the worries themselves'.¹⁹³ Williams objected to this on the grounds that if the resemblance of metaphysical concerns to neuroses were taken really seriously, the philosopher ought to hand over to the clinical psychoanalyst. If the philosopher had anything to contribute to the 'cure' of the tortured metaphysician, then it was by offering some evaluation of the arguments themselves. In this, his aims in philosophy are much closer to those of *The Concept of Mind* than they are to those of the *Philosophical Investigations*.

A yet more fundamental reservation of Williams's about the project of the ordinary language tradition was its mischaracterisation, as he had it, of the *ordinary*. The idea behind one therapeutic conception of philosophy, he said in a lecture delivered late in his life, was

that philosophy is alienated from everyday consciousness, its speech from what someone would say outside philosophy: and the question must be, what this itself is taken to be. What it should not be taken to be – and this is the most important point – is merely banal conversation. ... The idea that it should be understood in terms of banal conversation was a mistake made by some "ordinary language" philosophy, which identified ordinary language as the language of a life rendered ordinary by the subtraction of the imagination ...¹⁹⁴

¹⁹³ Williams, *Humanistic Discipline*, 22. See also Williams's review of a Festschrift for Wisdom, reprinted in Williams, *Essays and Reviews*, 101–4.

¹⁹⁴ Williams, *Humanistic Discipline*, 208. The subsequent paragraphs of the essay come close to exempting the work of Stanley Cavell from the criticism – chiefly in virtue of Cavell's attempt 'to engage with history or our present cultural situation, particularly through its involvement with modernism' (210). The point is deeply interesting but there is no way of doing justice to the comparison in brief. Cavell was one of the dedicatees of *Descartes: The Project of Pure Enquiry* (London: Penguin Books, 1978).

In this, his criticisms are close to those of another contemporary critic of the regnant conceptions of philosophy at Oxford, though one from Hare's generation, Iris Murdoch.¹⁹⁵ Like Williams, she too had had a classical education, and had great regard for Plato. She too had read Wittgenstein with care. Her other influences were more recherché: the theologian Simone Weil and the French existentialists, in particular Sartre, on whom she wrote one of the first serious monographs in English. She, like Williams, saw some of the virtues of the ordinary language tradition: 'It is certain a great merit of this tradition, and one which I would not wish to lose sight of, that it attacks every form of spurious unity'.¹⁹⁶ But she worried, as Williams would later join her in doing, that the tradition represented an '[u]nambitious optimism' characteristic of 'the Anglo-Saxon tradition; ... it is not surprising that a philosophy which analyses moral concepts on the basis of ordinary language should present a relaxed picture of a mediocre achievement.'¹⁹⁷

Murdoch's writings in the 1950s focused their attention on the pro-behaviourist strands in Ryle's thought, particularly those that denied to moral philosophy the picture of moral thought and sentiment that relied on some assumption of an inner

¹⁹⁵ This (again, deeply interesting) comparison too will have to await its elaboration in a different place.

¹⁹⁶ Iris Murdoch, *Existentialists and Mystics: Writings on Philosophy and Literature* (Penguin Books, 1999), 340. See also her remark (at 319) on her regard for Austin's *Sense and Sensibilia*.

¹⁹⁷ Murdoch, *Existentialists*, 340. The best, most extensive, treatment of her critique is the editor's introduction to Justin Broackes, ed., *Iris Murdoch, Philosopher* (Oxford: Oxford University Press, 2011).

mental life. Twinned with a too relentless focus on *ordinary* language, she found that the simple psychological descriptions to be found in Ryle and Hare were inadequate to the phenomena. A passage in her classic 1956 paper read at the Aristotelian Society, 'Vision and Choice in Morality' says:

There are situations which are obscure and people who are incomprehensible, and the moral agent ... may find himself unable to describe something which in some sense he apprehends. Language has limitations and there are moments when, if it is to serve us, it has to be used creatively, and the effort may fail. When we consider here the role of language in illuminating situations, how insufficient seems the notion of linguistic moral philosophy as the elaboration of the evaluative-descriptive formula. From here we may see that the task of moral philosophers has been to extend, as poets may extend, the limits of the language, and enable it to illuminate regions of reality which were formerly dark.¹⁹⁸

Her criticisms went further, and encompassed, in an inspired and unobvious feat of intellectual taxonomy, ideas she found to be shared by such Oxford philosophers as Ryle and Hare on the one hand and such French philosophers as Sartre on the other. In a lecture delivered in 1976, she presented the following intellectual historical narrative of her peers (in her words, 'of British philosophy since Moore'):

After puerile attempts to classify moral statements as ... expressions of emotion, a more sophisticated neo-Kantianism with a utilitarian atmosphere has been developed. ... [T]he agent ... appears as an isolated will operating with the concepts of 'ordinary language' ... Linguistic analysis claims simply to give a philosophical description of the human phenomenon of morality, without making any moral judgements. In fact the resulting picture of human conduct has a clear moral bias. The merits of linguistic analytical man are freedom (in the sense of detachment, rationality), responsibility, self-awareness, sincerity, and a lot of utilitarian common sense. There is of course no mention of sin, and no mention of love.¹⁹⁹

¹⁹⁸ Originally Iris Murdoch, "Vision and Choice in Morality," *Proceedings of the Aristotelian Society: Dreams and Self-Knowledge* Supplement no. 30 (1956): 32–58. The quotation is from Murdoch, *Existentialists*, 90.

¹⁹⁹ Murdoch, *Existentialists*, 339–40. Cf Williams, *Essays and Reviews*, 312: 'The aim ... of being ... concerned only with the form of moral thought and not with its content ... laid itself open to the

Murdoch's unorthodox vocabulary, her distinctive prose style and the command of varied registers evident in both her philosophy and in her novels, and her unOxonian breadth of reference, marked out the dissident path that Williams would take soon after. But there were important differences. For one thing Murdoch, while she never lost her interest in philosophy, did not stay in an academic position for the full length of her career; Williams by contrast had a long and by any conventional standards hugely successful career in mainstream academia. But the differences ran deeper. For all his injunctions to imaginativeness in philosophical prose, he was formally unadventurous and did not range far beyond the philosophical essay, lecture, or monograph as a vehicle for his thought. They shared an interest in metaphors of *perception* as improvements on previous ways of thinking about ethics – and on this point I shall have a great deal more to say – but Murdoch's distinctive vocabulary ('vision', 'attention', and as a result of her interests in Buddhism, 'unselfing') bespeak a temperament quite unlike Williams's. The most important difference between their ideas of ethical perception – and indeed much else – can be captured in the contrasting attitudes they took to Nietzsche.

charge of obstructing real moral discussion, because substantive moral assumptions were hidden by the methodology.' Also, Williams, *Sense of the Past*, 54: 'Kantianism and consequentialism, despite their other differences ... resemble each other, as Iris Murdoch has insisted, in sharing a concern with the practical. ... the situation of the rational agent intending to change the world preoccupies them both, and the very plain fact that everything that an agent most cares about typically comes from, and can be ruined by, uncontrollable necessity and chance is no part of their concerns.'

Murdoch was largely indifferent to the appeal of Nietzsche, much preferring Schopenhauer. Williams, however, was a convert (he balked of course at his politics).²⁰⁰ By the late 70s, he had abandoned the prejudice that had him ask a colleague in the early 1960s ‘Why do you waste time over rubbish that [the popular philosopher and broadcaster CEM] Joad could have refuted?’²⁰¹ He would declare in 1981:

It is certain, even if not everyone has yet come to see it, that Nietzsche was the greatest moral philosopher of the past century. This was, above all, because he saw how totally problematical morality, as understood over many centuries, has become, and how complex a reaction that fact, when fully understood, requires. To help himself to understand it, he resourcefully explored, in twenty years of increasingly hectic activity, our feelings about art, guilt, violence, honesty, and indeed every element of that moral consciousness which the Greeks helped to invent.²⁰²

A philosopher who found himself sympathetic to such a vision of morality had to part company with anyone with a more comforting diagnosis. In 1977, he would, in a searching, often admiring, but ultimately negative, review of Murdoch’s attempt in *The Fire and the Sun* to defend a Platonic ethic that identified truth, reality and

²⁰⁰ See, e.g., Williams, *Shame and Necessity*, 10: ‘[Nietzsche] did not move to any view that offered a coherent politics. He himself provides no way of relating his ethical and psychological insights to an intelligible account of modern society—a failing only thinly concealed by the impression he gives of having thoughts about modern politics that are determinate but terrible.’

²⁰¹ Jane O’Grady, “Professor Sir Bernard Williams,” *The Guardian*, June 13, 2003.

²⁰² Williams, *Essays and Reviews*, 183. Laudatory remarks on Nietzsche as a moral philosopher appear in, e.g., Bernard Williams, *Making Sense of Humanity and Other Philosophical Papers 1982–1993* (Cambridge: Cambridge University Press, 1995), 65–76; Williams, *Shame and Necessity*, 9 and *passim*; Williams, *Truth and Truthfulness*, 12–13 and *passim*.

goodness, and a picture of ‘morality in terms of abandoning ego in favour of reason, reality, or duty’:

Suppose we ... rest the weight on a conception of the good as something external to the ‘greedy’ ego ..., something inherent in a hard order or reality. What will we or [Murdoch] say to the claim ... that there is no such order of things, that what she would have us yearn for is not there? This is an embarrassment which has been latent in other of Miss Murdoch’s writings, and not solely in her philosophy. Faced with the historical retreat from a conception of the moral order as transcendental, one who nevertheless insists that goodness lies in correct perception and freedom from error will inevitably tend to locate the subject-matter of that knowledge or error merely in the world around us: the misperception which is sin will have to consist in a misunderstanding not of the cosmic order of things, but (for instance) of other people. In that direction, in total and ironical contrast to the original spirit of the thing, lies cosiness. Only connect, understand others, you cannot will evil. It is a pity ... that Nietzsche has not yet succeeded in persuading us what a hopeless thought that is.²⁰³

These are sharp words, but this is not the place to consider whether they are fair.²⁰⁴

I quote at length to mark the gap that Williams’s emerging Nietzschean sympathies would open between him and other critics of the sorts of moral philosophy he scorned. I shall briefly draw one last contrast before I turn to Williams’s alternative to the orthodoxy. Some of Williams’s discontents, certainly some of those registered in his semi-polemical Preface to *Morality*, suggest concerns that Elizabeth Anscombe had voiced some years earlier. Her astringent and in places intemperate 1958 article ‘Modern Moral Philosophy’ stands out as an early, and courageous, stand against the

²⁰³ Williams, *Essays and Reviews*, 144–5.

²⁰⁴ Williams, *Truth and Truthfulness*, 199–205 has a section (‘Authenticity and other people’) that might well be an attempt to tell find a way of making room for Murdoch’s idea that ethics was all about the acknowledgement of the reality of other people, without incurring either Nietzsche’s scepticism or his own charge against Murdoch of ‘cosiness’. In fact, the chapter in *Truth and Truthfulness* on the virtue of ‘Accuracy’, owes an evident, if unacknowledged, debt to Murdoch.

dominant trends of the discipline at the time.²⁰⁵ The article remains a rich source of provocation – in its denunciation of ‘consequentialism’ (an Anscombe coinage) for the moral perversities it licensed, in its accent on the relation between such notions as ‘moral duty’ and their historical origins in and dependence on the foundations provided by theological discourse, and most importantly of all, its recommendation of a moratorium on moral philosophy until a more adequate moral psychology had been developed. That last recommendation is thought to have inaugurated the tradition sometimes labelled ‘virtue ethics’.²⁰⁶

Williams had great regard for Anscombe, in particular for the seriousness she brought to an intellectual milieu characterised by its reverence for cleverness.²⁰⁷ They were separated by their contrasting attitudes to the legacy of Wittgenstein, Aristotle, and perhaps most importantly of all, Christianity. This was most apparent in an essay Williams co-authored, with Michael Tanner, replying to an article of

²⁰⁵ G. E. M. Anscombe, “Modern Moral Philosophy,” *Philosophy* 33, no. 124 (1958): 1–19. See Timothy Chappell, “Review: Bernard Williams, Philosophy as a Humanistic Discipline, The Sense of the Past,” *Philosophical Investigations* 32, no. 4 (October 2009): 362–3 for a discussion of the absence of references to Anscombe in Williams’s work.

²⁰⁶ E.g. ‘... Anscombe’s challenge has been taken up by many philosophers, most of them intent on developing or meeting it. One effect of this work has been the emergence, or perhaps re-emergence, of what is now called “virtue ethics”’ (Roger Crisp, “Modern Moral Philosophy and the Virtues,” in *How Should One Live?: Essays on the Virtues*, ed. Roger Crisp (Oxford: Oxford University Press, 1998), 2).

²⁰⁷ Alex Voorhoeve, *Conversations on Ethics* (Oxford: Oxford University Press, 2009), 197 quote Williams saying: ‘Oxford philosophy tended to be clever, but Elizabeth [Anscombe] conveyed a strong sense of the seriousness of the subject, and how the subject was difficult in ways that simply being clever wasn’t going to get round’.

Anscombe's that defended the Catholic Church's position on contraception.

Williams and Tanner claimed that her attempt to defend the distinction between the (permissible) 'rhythm' method of contraception and other (impermissible) methods was unsustainable, and made a case for the liberal sexual morality against which Anscombe had set herself.²⁰⁸ Williams's later work would show considerable interest in the questions of moral psychology and the philosophy of action that Anscombe had done so much to make respectable, but he came at the questions in a spirit quite opposed to hers. Williams the lifelong atheist may well have said of Anscombe's God what he said about Murdoch's (and Plato's) Good: 'what she would have us yearn for is not there'.

3. Liberation philosophy?

A brief recap might be called for at this point. We have seen that Williams had some respect for the ideal of (moral) seriousness and the approach to the philosophy of action that Anscombe represented, but did not share her regard either for Wittgensteinian methods or for the Catholic Church. He shared with Murdoch an interest in a moral philosophy tied to a richer (non-behaviourist) picture of moral psychology but found her Platonic aspirations neither attractive nor sustainable. He

²⁰⁸ Bernard Williams and Michael Tanner, "Comment on 'Contraception and Chastity,'" *The Human World*, 1972.

shared with Austin and the later Wittgenstein their dislike of essentialism and their focus on the distinctiveness of the individual case, but rejected the idea of a linguistic examination detached from theoretical concerns that might give the assembly of linguistic data its point. Further, he rejected any therapeutic conception of philosophy that divorced therapeutic concerns from a specifically *historical* diagnosis of the need for therapy.

All of this sits well with the self-styled Nietzscheanism of the second half of his career. What is surprising is that of all his parochial, that is to say Oxonian, influences, it is with Ryle's project in *The Concept of Mind* that Williams's work had the greatest affinity – or so I shall argue in this section. The two philosophers' projects may both be characterised as species of (what I shall venture to call) *liberation* philosophy: argument aimed at helping the philosopher, indeed any reflective person, shed certain inherited dogmas that distort perception of the relevant phenomena. The comparison with Ryle, of all people, might strike a reader as strange. Williams had no time for any sort of behaviourism,²⁰⁹ and he found much in Ryle's later work to reject; in general, as I shall argue, Williams was convinced of the reality and complexity of the 'inner life' and its relevance to philosophy, particularly moral philosophy. The connection to Ryle lay at a more basic, we might

²⁰⁹ His sharpest, and most amusing, remarks against behaviourism appear in a review of B. F. Skinner (Williams, *Essays and Reviews*, 87–9).

say *structural*, level, and might be stated in one sentence: Williams's arguments against what he came to call the 'morality system' have the same structure as Ryle's arguments against the 'Cartesian' view of the mind.

A précis of some of the similarities I shall highlight in this section: In both cases, the systems in question are ubiquitous in philosophy but not entirely the inventions of philosophers. They are both distortions, or mischaracterisations, of something real in our experience, and have been given their most abstract expression in the writings of philosophers (namely, Descartes and Kant). In both cases, there are powerful *historical* factors that explain the appeal of the erroneous doctrine. In both cases, philosophical argument – as opposed, say, to psychoanalysis – might have at least some traction on the misconception and act as a prophylactic against any future temptation to fall into it. And happily, a better philosophy might be able to do justice to the phenomena and practices that the misconceived doctrine took for its material and offer a manner of liberation from the distortions of the doctrine.

The first chapter of *The Concept of Mind* opens with these words:

There is a doctrine about the nature and place of minds which is so prevalent among theorists and even among laymen that it deserves to be described as the official theory. Most philosophers, psychologists and religious teachers subscribe, with minor reservations, to its main articles and, although they admit certain theoretical difficulties in it, they tend to assume that these can be overcome without serious modifications being made to the architecture of the theory. It will be argued here that the central principles of the doctrine

are unsound and conflict with the whole body of what we know about minds when we are not speculating about them.²¹⁰

The doctrine ('which hails chiefly from Descartes') then gets Ryle's famous label: 'the dogma of the Ghost in the Machine'.²¹¹ An intriguing, and not altogether disreputable, genealogy is then found for it. The dogma turns out to be an inventive reaction to the dilemma of a supremely intelligent religious man impressed by the scientific claims of post-Galilean mechanics but unwilling to risk the impiety of embracing a mechanical model of the human mind. A later passage in Ryle complicates this historical picture somewhat. It turns out that the ground for a Cartesian philosophy of mind were laid much earlier – the theology of both mediaeval scholasticism and the Reformation, themselves shaped by a 'Stoic-Augustinian' philosophy of action that was itself grounded in the basic details of what was common to the psychology of Plato and Aristotle. 'Descartes was reformulating already prevalent theological doctrines of the soul in the new syntax of Galileo. The theologian's privacy of conscience became the philosopher's privacy of consciousness'.²¹²

²¹⁰ Ryle, *Concept of Mind*, 1.

²¹¹ Ryle, *Concept of Mind*, 5.

²¹² Ryle, *Concept of Mind*, 13.

This is not the place to ask how much there is much to be said for this story; it is enough to say it is not absurd, and that Ryle was not the only person to conjecture this sort of explanation.²¹³ What is striking is the fact that Ryle thought it important to offer a historical explanation for the currency of the dogma. (Austin's *Sense and Sensibilia* teases its reader with the hint of some such story about how the notion of 'sense-data' ever got to its position of prominence among philosophers of the mid-20th century, but the story never comes.²¹⁴) In this, and not only in this, *The Concept of Mind* would have been likelier than anything by Austin or Wittgenstein or even Cavell to appeal to Williams's evolving Nietzschean affinities.²¹⁵

Consider now these early sentences from the final chapter of Williams's *Ethics and the Limits of Philosophy*, titled 'Morality, the Peculiar Institution'. The title is a provocation: 'the peculiar institution' was the euphemism the Confederacy in the American Civil War used for slavery. The reference to 'morality', by this stage in the

²¹³ Williams had a subtler version of a similar historical claim: see Bernard Williams, *Descartes*, 276–7 and *passim*.

²¹⁴ Austin, *Sense and Sensibilia*, 61: 'the apparent sophistication of Ayer's "linguistic" doctrine really rests squarely on the old Berkeleyian, Kantian ontology of the "sensible manifold". He has all along, it seems, really been completely convinced by the very arguments that he purports to "evaluate" with so much detachment. And there can be little doubt that this is owing in large measure to his wholesale acceptance of the traditional, time-hallowed, and disastrous manner of expounding them'; see also *Sense and Sensibilia*, 104: 'The pursuit of the incorrigible ... is rampant all over ancient philosophy, most conspicuously in Plato, was powerfully re-animated by Descartes, and bequeathed by him to a long line of successors. No doubt it has many motives and takes many forms, and naturally we can't go into the whole story now.'

²¹⁵ Again, see Williams, *Humanistic Discipline*, 208 for a brief discussion of his qualified sympathy for the (relative) historical sensitivity of Cavell's Wittgenstein, and an all-too-brief hint at his discontents with the 'Heideggerian resonances' of even Cavell's picture of modernity.

book's argument, is not a simple description of its subject matter. Morality, by this stage in Williams's career, stands in contrast to – and is not identical to – *ethics*, the most general set of inquiries into the Socratic question of 'how one should live', as Williams describes the project in the opening chapter of the book.²¹⁶ Morality, he says, is 'a particular development of the ethical', characterised by an emphasis on 'a special notion of obligation'.²¹⁷ It is only in the book's final chapter that he spells out what he takes the peculiarities of this answer to the Socratic question to be:

The important thing about morality is its spirit, its underlying aims, and the general picture of ethical life it implies. ... Morality is not one determinate set of ethical thoughts. It embraces a range of ethical outlooks ... The philosopher who has given the purest, deepest, and most thorough representation of morality is Kant. But morality is not an invention of philosophers. *It is the outlook, or, incoherently, part of the outlook, of almost all of us.*²¹⁸

There is every now and then in Williams the urge to fall into something like Ryle's position as regards the Cartesian dogma, namely to claim that 'the central principles of the doctrine ... conflict with the whole body of what we know about minds when we are not speculating about them'.²¹⁹ There is a faint echo of this in his cynically self-deprecating remark about his work having 'consisted largely of reminding moral philosophers of truths about human life which are very well

²¹⁶ Williams, *Ethics*, 4–5.

²¹⁷ Williams, *Ethics*, 6.

²¹⁸ Williams, *Ethics*, 174; emphasis added. Cf the closing lines of the chapter: 'Its [sc morality's] philosophical errors are only the most abstract expressions of a deeply rooted and still powerful misconception of life' (196). It is not clear whether his 'almost' means that he is excepting himself from the generalisation; certainly he admitted to an interviewer that he once 'used to have very pious utilitarian views. But I came to see that consequentialist reasoning could just lead you on and on in the wrong direction' (Jeffries, "The Quest for Truth").

²¹⁹ Ryle, *Concept of Mind*, 1, quoted above.

known to virtually all adult human beings except moral philosophers'.²²⁰ It should not be forgotten – and Williams is certainly anxious in narrating the story to tell us – that the remark was the product of a melancholy evening's drunkenness. Indeed the remark is at least in tension with some other things he said, including the remark above. If there was something of which moral philosopher stood in need of being reminded, it was not something that only philosophers had forgotten, and not something a chat with a folksy stranger on the Clapham omnibus would correct.

Certainly, turning the focus of philosophy from 'thin' concepts to 'thick' ones would not solve the problem – whatever his earlier critique of Hare's project might have seemed to suggest. For one thing, as a 1990 essay in the *Times Literary Supplement* had it, if one restricts the objects of philosophical analysis to the concepts immanent in the practice of the community to which one belongs, it will be difficult to find a basis either for criticism of those concepts and practices, or indeed even for explaining how these practices could come to change, unless – paradoxically – they had already changed.

A philosophical account that considers only the concepts that we pick up from our local community will find it hard to explain the criticism and alteration of ethical practices. It runs the risk of sharing with a certain kind of Right Hegelianism ... a cultivation of an inarticulate conservatism of the folk-ways. It is natural that people who are anyway drawn to a Right Hegelian enthusiasm for the folk-ways (or rather, as is their habit, for a condescending fantasy about the folk-ways) often welcome the Wittgensteinian line,

²²⁰ Williams, *In the Beginning*, 52.

interpreted in this way. They see it as the continuation of Hegelian conservatism by other means.²²¹

The echo with Von Clausewitz ('War is the continuation of politics by other means') is precisely well chosen. What was needed, he said, was a '*left-Wittgensteinianism*', that is to say 'a view that accepts the insights about the thickness of our primary ethical understanding and its relation to social practices, but leaves room for a radical critique in the name of interests not adequately expressed in the folk-ways'.²²² The idea, unlikely to give its name to a popular political movement, was developed in another essay published two years later, and elicited from Williams a characteristic aphorism: 'Practice is not just the practice of practice, ... but also the practice of criticism'.²²³ He continued:

Once we regard the ethical life we now have as a genuinely historical and local structure, one that is peculiarly self-conscious about its own origins and potentialities, we shall have less temptation to assume that it is a satisfactorily functioning whole; and we shall be more likely to recognize that some widely accepted parts of it may stand condemned in the light of perfectly plausible extrapolations of other parts.²²⁴

If one adopts a historicist attitude to current ethical practices, it will be natural to look in them for the irregularities characteristic of historically contingent practices – as one does in studying the grammar of natural languages, for instance.

Philosophy done in this spirit could offer a picture of ethical thought that opened up the 'possibility of deploying some parts of it against others, and of reinterpreting

²²¹ Williams, *Essays and Reviews*, 315.

²²² Williams, *Essays and Reviews*, 315.

²²³ Williams, *In the Beginning*, 36.

²²⁴ Williams, *In the Beginning*, 36–7.

what is ethically significant, so as to give a critique of existing institutions, conceptions, prejudices, and powers'.²²⁵

Williams did not come by his ideas on these questions in a single epiphany; we can trace their evolution in his thought over the course of a long career. The idea that there is something outside of 'morality', narrowly construed, half-present in our consciousness, is only incipient in Williams's earliest work. It is given a preliminary development in his critique of utilitarianism for its implausible ideas about human agency (a critique I shall discuss below). It is developed further in his treatment of the question of 'moral luck' and the ineliminable character of certain moral sentiments. The idea is full-blown in *Ethics and the Limits of Philosophy*, but the historical story remains sketchy. Williams finally attempts an articulation of the historical origins of 'morality' in the lectures on Greek tragedy that make up *Shame and Necessity*, and his historicism is further developed in the many fine essays he wrote in the 1990s, culminating in *Truth and Truthfulness*, the last of his books to be published in his lifetime, where we find a genuinely constructive alternative to the morality system that seeks to be true to that system's raw material but to stay clear of the fallacies built into its nature.

²²⁵ Williams, *In the Beginning*, 36–7.

That brisk survey deserved an elaboration. One of the earliest appearances of such thoughts is in a lyrical passage of his first book, *Morality*, where Williams commends ‘a phrase of D. H. Lawrence’s in his splendid commentary on the complacent moral utterances of Benjamin Franklin: “Find your deepest impulse, and follow that.”’ The context is the question of whether morality (the term has not yet contracted to the narrow sense it acquires in Williams’s later works) is necessarily, constitutively, concerned with human well-being. The chapter concludes:

... if someone said – obscurely enough – that men need a world in which there is risk, uncertainty, and the possibility of despair, then a morality which emphasized this, as opposed to moralities which want as much as possible tidied up, might still be said to be concerned with men’s well-being. Something will still be excluded by the use of this term: systems of values or precepts which paid no attention at all to what we can understand men as needing or wanting.²²⁶

Something of this Lawrentian spirit can be found in his use of the notion of integrity as part of his critique of utilitarianism. I shall have more to say in a subsequent section about the two well-known examples (‘George’ and ‘Jim’) he uses to make his case. Consider for now Williams’s remark about what was omitted from purely utilitarian ways of looking at certain sorts of difficult ethical decisions that involve some obscure tension with non-utilitarian sentiments we might have:

to reach a grounded decision in such a case should not be regarded as a matter of just discounting one’s reactions, impulses and deeply held projects in the face of the pattern of utilities, nor yet merely adding them in – but in the first instance of trying to understand them.²²⁷

²²⁶ Williams, *Morality*, 95.

²²⁷ Williams, “Critique”, 118.

Utilitarianism, by the end of the essay, stands charged with chronic simple-mindedness, a condition that ‘consists in having *too few thoughts and feelings* to match the world as it really is’. Nothing in that statement implies that something more primal, call it ‘common-sense morality’, does a better job; but whether our non-utilitarian moral sentiments are or are not adequate to the world as it really is becomes a more urgent question in the face of utilitarianism’s simple-mindedness, ‘the number of dimensions in which it runs against the complexities of moral thought’.²²⁸ The effort, ‘in the first instance of trying to understand them’ – that is to say, to understand those complexities – is a philosophical challenge, and it is one that Williams’s future work took on as its central task.

The essays published in the years after ‘A Critique of Utilitarianism’ were collected in 1981 as *Moral Luck*, a volume that opened with the prefatory reflection on his ‘widening doubt’ about morality, ‘which has led me to try to find out – often by the crude method of prodding it – which parts of moral thought seemed to me to be actually alive, before trying to design any elegant physiology for it’.²²⁹ The title essay was the clearest example of the ‘prodding’ method in practice. His procedure, he declared early in the essay, would be to ‘invite reflection about how to think and feel’ about certain unusual situations by inviting reflection on more usual ones. This

²²⁸ Williams, “Critique”, 149.

²²⁹ Williams, *Moral Luck*, x.

was not by the device that has come to be known as ‘intuition pumping’; Williams was interested rather in ‘the *experience* of those kinds of situations’. He was going to describe a certain pattern of thought and sentiment that was ‘possible, coherent, and intelligible’, perhaps even rational. To live without these sentiments, he said, ‘would involve a much vaster reconstruction of our sentiments and our view of ourselves than may be supposed’.²³⁰

This was rather like the argument of ‘A Critique of Utilitarianism’, and the similarity was not only methodological. Again the target was a certain idea of morality – declared by some philosophers to be the only rational one – that involved, as a little reflection would reveal, some quite startling disruptions to common, and deeply held, assumptions about the relationship between a person and her actions. The essay offered examples that have come to be famous in moral philosophy – the case of ‘Gauguin’ (both like and unlike the actual painter) and Tolstoy’s Anna Karenina, but also that of a nameless

lorry-driver who, through no fault of his, runs over a child[.] [This driver] will feel differently from any spectator ... Doubtless, and rightly, people will try, in comforting him, to move the driver from this state of feeling, move him indeed from where he is to something more like the place of a spectator, but it is important that this is seen as something that should need to be done, and indeed some doubt would be felt about a driver who too blandly or readily moved to that position. We feel sorry for the driver, but that sentiment co-exists with, indeed presupposes, that there is something special about his

²³⁰ Williams, *Moral Luck*, 22.

relation to this happening, something which cannot merely be eliminated by the consideration that it was not his fault.²³¹

The agent would feel a special variety of regret, characterised by more than the thought ‘how much better if it had been otherwise’. This ‘agent-regret’ – a coinage of Williams’s²³² – is something ‘a person can feel only towards his own past actions’,²³³ and expresses itself in a set of distinctive attitudes and sentiments, and often in certain kinds of action (as when the *ex hypothesi* innocent lorry-driver might feel moved to make some gesture to the family of the victim).

... that there is room in the area for irrational and self-punitive excess, no one is likely to deny. But equally it would be a kind of insanity never to experience sentiments of this kind towards anyone, and it would be an insane concept of rationality which insisted that a rational person never would. To insist on such a conception of rationality, moreover, would, apart from other kinds of absurdity, suggest a large falsehood: that we might, if we conducted ourselves clear-headedly enough, entirely detach ourselves from the unintentional aspects of our actions ... and yet still retain our identity and character as agents.²³⁴

The ‘widening doubt’ in *Moral Luck* about morality and the ideals of rationality it embodied was the central theme of *Ethics and the Limits of Philosophy*, and the target was now the consequences of this doubt for the project often thought to be constitutive of moral philosophy: the quest for a moral theory that could enable us to revise practices and beliefs that are based on mere ‘prejudice’. Williams declared that such an aspiration came from ‘the assumption that we probably have too many

²³¹ Williams, *Moral Luck*, 28.

²³² It is significant about his views on ‘morality’, and relevant to our discussion of his relationship with ordinary language philosophy, that Williams preferred this clunky neologism to ‘remorse’.

²³³ Williams, *Moral Luck*, 27.

²³⁴ Williams, *Moral Luck*, 29–30.

ethical ideas, some of which may well turn out to be mere prejudices. Our major problem now is actually that we have not too many but too few, and we need to cherish as many as we can'.²³⁵

His next book, *Shame and Necessity*, was a scholarly study of some themes in Homer and Greek tragedy that was also a development of themes in his moral philosophy. At the risk of being thought a 'classicizing reactionary',²³⁶ he wanted to argue that it was possible to recover from extant ancient Greek texts an ethical outlook that was in important respects like our own, and to the extent that it differed, it was we who came out the worse in the comparison. 'It is not,' he said, 'a question of *reviving* anything. What is dead is dead ... What is alive from the Greek world is already alive and is helping (often in hidden ways) to keep us alive.'²³⁷

Against interpreters of Homer and other Greeks who claimed that all subsequent western civilization involved progress from their primitive notions of action, will and responsibility, Williams wanted to urge that the Greeks had most everything

²³⁵ Williams, *Ethics*, 117. Cf Williams, "Critique", 149, quoted above, on utilitarianism having 'too few thoughts and feelings to match the world as it really is'. Cf also the celebrated sentence in 'Persons, Character and Morality' about the man whose motivating thought in saving his wife rather than a stranger from a fire was not 'the thought that it was his wife' but 'that it was his wife and that in situations of this kind it is permissible to save one's wife': such a man would have 'one thought too many' (Williams, *Moral Luck*, 18). In Williams's ethics, as in Aristotle's, the mean is hard to hit.

²³⁶ Williams, *Shame and Necessity*, 6.

²³⁷ Williams, *Shame and Necessity*, 7.

they needed from such ideas; what was purportedly missing from their worldview was an artefact of bad philosophy, of the sort that received its purest exposition in Kant. The idea, in particular, that there was a part of human life that could be wholly immune to luck was one that – for all the appeal of its vision of perfect justice – the Greeks lacked, and rightly so.²³⁸

A newspaper interview shortly after the publication of *Shame and Necessity* had him face, again, the charge that his writings were entirely negative. He replied, with a directness rare in his writings:

I don't see [my work] as negative, I see it hopefully as liberating. It seems to me people get themselves in situations in which they feel they have no right to have certain kinds of moral thoughts because they don't fit in with ... some very impoverished theoretical picture of what constitutes moral thought. Roughly, if it isn't about obligation or consequences, it doesn't count. That's not the way most people think most of the time about most things.²³⁹

The echo of Ryle ('the central principles of the doctrine ... conflict with ... what we know about minds when we are not speculating about them') are obvious. Williams put this point more precisely in a reply to his critics:

If philosophical destruction of theories in ethics can provide a liberation, it had better do more than liberate philosophy from those theories. Its hope, rather, must be to liberate a reader, someone who wants to think about such questions, from distortions or misunderstandings involved in his or her own experience, and the aim of this (it applies as much to the writer as the reader) is the Socratic aim of improving that experience. The

²³⁸ See also Williams, *Ethics*, 196: 'The idea of a value that lies beyond all luck is an illusion.' A judicious recent discussion of Williams's revisionist historiography of archaic Greece is Roger Crisp, "Homeric Ethics," in *The Oxford Handbook of the History of Ethics*, ed. Roger Crisp (Oxford: Oxford University Press, 2013).

²³⁹ John Davies, "A Fugitive from the Pigeonhole," *Times Higher Education*, November 4, 1996.

destruction of theory is able to play a role in this because a powerful philosophical theory can be an effectively articulated expression of those distortions. ... [P]hilosophical theory may be one expression of misconceptions in our ethical consciousness, which can be exposed to some extent by philosophy ... but are not to be explained primarily by a mysterious power of philosophical theory to invade healthy practice. They are explained, like our other personal and cultural discontents, by many kinds of considerations, including, no doubt, some other kinds of theory.²⁴⁰

But Williams's thoughts on the question went deeper, and they were sensitive, as Ryle was not, to the question of who the 'we' in question were. A helpful endnote in *Shame and Necessity* spelt out to whom he was referring with his ubiquitous first-person plural pronouns:

The best I can say is that "we" operates not through a previously fixed designation, but through invitation. ... It is not a matter of "I" telling "you" what I and others think, but of my asking you to consider to what extent you and I think some things and perhaps need to think others.²⁴¹

This point is not always appreciated by Williams's readers. A severe, and in many ways representative, misunderstanding of his point appears in a (for the most part laudatory) review by Jonathan Barnes of *Shame and Necessity* for the *Times Literary Supplement*.²⁴² Barnes writes that '[f]or the most part, ... "we" seem to be down-at-heel moral philosophers of a vaguely Kantian bent.' Elsewhere he writes that

²⁴⁰ Altham and Harrison eds., *World, Mind, Ethics*, 218–9. A small point may be noted regarding Williams's use of 'his or her own experience': it is only in the writings of the 1990s that Williams, like many others of his generation, consistently eschews the use of masculine pronouns for hypothetical persons and 'man' for 'human being', effectively an acknowledgement of the force of the feminist claim about how presuppositions (in this case embodied in the conventions of prose writing) can distort experience.

²⁴¹ Williams, *Shame and Necessity*, 171 n. 7. On the designation of 'we', see also Williams, *Humanistic Discipline*, 196 and Williams, *Moral Luck*, 159–61.

²⁴² Jonathan Barnes, "Like Us, Only Better," *The Times Literary Supplement*, no. 4699 (April 23, 1993): 3–4.

“Williams’s opponents ... are not numerous; and they are all philosophers, or men marked by the stigma of philosophy. It is philosophy – moral philosophy, Kantian moral philosophy – which here bears the responsibility and the guilt for “our” degenerate ethical attitudes.’ But despite claiming to have ‘little taste for tripe à la mode de Kant’, Barnes then shows evidence of a quite strikingly Kantian sensibility in his curt response to Williams’s remarks about the inevitability and aptness of agent-regret even in situations of blameless wrong-doing.²⁴³

Well, I hope that I should have no such feelings myself, and I do not admire people who do have them. ... perhaps I ought to feel an especial horror that I was myself involved in it [some terrible event]; but to feel guilt and to claim responsibility [sic: Williams speaks only of feeling ‘agent-regret’] seem to testify to an intellectual muddle rather than to a refined moral sensibility.²⁴⁴

Of course, this rather begs the question against Williams, who was arguing against precisely the idea that this must be an intellectual muddle. It is also unclear what exactly is involved in the ‘especial horror’ Barnes allows that one ought to feel in such circumstances and why that would not testify to the same intellectual muddle. Consider again the driver who, awake and sober, ran over a child who ran out on the street too unexpectedly for him to brake in time. He is likely to feel a connection with the incident that survives his friends’ quite correct assurances that there was nothing he could have done to prevent it – expressed by his wanting to make (e.g.) some symbolic gesture to the family of the child, given that no real restitution is

²⁴³ Williams, *Shame and Necessity*, 92–3, referring back to Williams, *Moral Luck*, 28–9 quoted above.

²⁴⁴ Barnes, “Like Us”.

possible. By stipulation, the lorry driver is not at fault; he was not negligent in any way. The point is that he is, nevertheless, the person who ran over a child, and this brute fact is likely to make some claim on his sentiments, and not one that is easily erased.²⁴⁵ In this, the lorry-driver is in a position not unlike that of Sophocles'

Oedipus:

The whole of the *Oedipus Tyrannus*, that dreadful machine, moves to the discovery of just one thing, that *he did it* [sc. that he, Oedipus killed his father]. Do we understand the terror of that discovery only because we residually share magical beliefs in blood-guilt, or archaic notions of responsibility? Certainly not: we understand it because we know that in the story of one's life there is an authority exercised by what one has done, and not merely by what one has intentionally done.²⁴⁶

'There is an authority exercised by what one has done', he says, and there is nothing irrational about submitting to this authority. This is not a matter of returning to common-sense morality. We are dealing with circumstances where the needs of reflection will not be satisfied by the half-articulated ideas in common-sense morality. The relevant category is not common sense but *experience*, and that experience needs to be made sense of in a way that does justice to the features of it that are apt to puzzle us. The task of making sense here is, among other things, a

²⁴⁵ For an articulate, affecting memoir by someone who was in exactly this position, see Darin Strauss, *Half a Life*, (New York: Random House Trade, 2011). Strauss ran over a bicyclist at the age of 18, but the incident never ceased to haunt him, even when the legal system exonerated him completely. The first line of the book – 'Half my life ago, I killed a girl' – captures exactly what Williams means when he writes that 'if one attaches importance to the sense of what one is in terms of what one has done ..., one must accept much that makes its claim on that sense solely in virtue of its being actual' (Williams, *Moral Luck*, 30).

²⁴⁶ Williams, *Shame and Necessity*, 69.

stylistic question: in what way can a piece of philosophical prose give an adequate account of ethical experience?

In his preface to a 1993 re-issue of his first book *Morality*, Williams returned to this old concern:

Sometimes literature or history can be called upon, to give some idea of the weight or substance of ethical concepts that we use or have been used by others; analytic argument, the philosopher's speciality, can certainly play a part in sharpening perception. But *the aim is to sharpen perception, to make one more acutely and honestly aware of what one is saying, thinking and feeling*. Philosophy invites us ... to ask whether what we say in morality is true. ... [I]t is vital not to forget another question that is to be asked both about morality and about moral philosophy, how far what we say rings true.²⁴⁷

'How far does what we say *ring* true?' 'The aim is to sharpen perception.' One can hear in both these phrases an echo of the voices of Williams's senior colleagues, Austin's fine linguistic 'ear' on the one hand and Murdoch's 'vision' on the other. But it is, unsurprisingly, the Nietzsche of the *Genealogy of Morality* with whom Williams seems to have the greatest affinity, the Nietzsche who describes so well our inattentiveness to our experiences, and our retrospective tendency to 'rub our ears and ask, astonished, taken aback, "What did we actually experience then?"'²⁴⁸

²⁴⁷ Bernard Williams, *Morality: An Introduction to Ethics*, Reissue (Canto, 1993), xiv–xv.

²⁴⁸ Friedrich Nietzsche, "*On the Genealogy of Morality*" and *Other Writings*, ed. Keith Ansell-Pearson, trans. Carol Diethe, 2d edition (Cambridge: Cambridge University Press, 2006), 3. The full quotation, well worth reading with our discussion of Williams in mind, is: 'As far as the rest of life is concerned, the so-called "experiences", – who of us ever has enough seriousness for them? or enough time? I fear we have never really been "with it" in such matters: our heart is simply not in it – and not even our ear! On the contrary, like somebody divinely absent-minded and sunk in his own thoughts who, the twelve strokes of midday having just boomed into his ears, wakes with a start and wonders "What hour struck?", sometimes we, too, afterwards rub our ears and ask, astonished, taken aback,

In this, Nietzsche is very much in the tradition of modern philosophy Kant had inaugurated. It is worth recalling that for all that his corpus might be seen as an extended attempt to refute Kant's moral philosophy, Williams did think that Kant was 'the greatest modern philosopher, though I don't agree with him.'²⁴⁹ In a late paper his death did not allow him to revise for publication, he found even in Kant a basically sound approach to ethics, aimed in the first instance at understanding our ethical experience:

Kant's moral philosophy certainly does not overlook everyday human experience; rather, it radically distorts it, not just by taking part, and a rather local part, of moral experience [viz. the phenomenology of obligation] as its essence but by doing so in the spirit of making final and vindictory sense of it, so turning us away from other kinds of sense we need to make of many other things that are more recalcitrant. ... [Still,] Kant's thought ... is fed by real features of moral experience and by demands and hopes, dimly felt, for that experience to be coherent and honourable in the conditions of modernity, and it is truthfully fashioned from those materials.²⁵⁰

The idea is that Kant was right to begin his moral philosophy in a critical examination of elements of moral experience, but his central error had consisted in his taking the experience of moral obligation as fundamental to that experience in a way that his arguments did not justify. Williams seemed to have a higher regard for Kant's theoretical philosophy:

"What did we actually experience then?" or even, "Who are we, in fact?" and afterwards, as I said, we count all twelve reverberating strokes of our experience, of our life, of our being – oh! and lose count . . . We remain strange to ourselves out of necessity, we do not understand ourselves, we must confusedly mistake who we are ...'

²⁴⁹ Bernard Williams, "Seminar with Bernard Williams," *Ethical Perspectives* 6, no. 3 (1999), 244.

²⁵⁰ Bernard Williams, "Fictions, Philosophy, and Truth," *Profession*, January 1, 2003, 39–40. This idea is helpfully discussed in A. W. Moore, *Noble in Reason, Infinite in Faculty: Themes and Variations in Kant's Moral and Religious Philosophy* (Routledge, 2003), 135.

Kant's attempt to make philosophy start from a reflection on the powers and limitations of observers such as ourselves. When Kant introduced this philosophy, he called it the *critical* philosophy. The process of critique that Kant initiated subsequently undid much of his own philosophy... It is hard to go on from here as though none of this had happened. We should not forget that the style of philosophy to which Kant self-consciously opposed his critique he called *dogmatic* philosophy, meaning that it took the supposed deliverances of reason at their face value, without asking how they were grounded in the structure of human thought and experience.²⁵¹

It seems that Williams conceived of his own work as critical in this Kantian sense, even if he was on the side of those concerned to 'undo' much of what Kant himself did under this flag. In his vision of ethics, even if philosophy cannot give the kind of help one might have expected of it – a theory to guide action – it might help, as A. W. Moore has precisely put it, to 'assist the self-understanding of those whose ethical reasoning already has guidance from elsewhere'.²⁵² From where, one might ask. To this, a short answer may be found in the oracular epigraph from Albert Camus to *Ethics and the Limits of Philosophy*: 'Quand on n'a pas de caractère, il faut bien se donner une méthode' (roughly, 'if one hasn't got a character, one had better get oneself a method').²⁵³ No one comes to the business of ethical reflection who is

²⁵¹ Williams, *Essays and Reviews*, 379 (the essay is a long review of Thomas Nagel's *The Last Word* in *The New York Review of Books*). Williams returns to the theme of this quotation later in the review: 'If we come to understand historically and psychologically how our own and others' ethical thoughts came about, this can change the way we think about the status of our thoughts, and about their relation to other people's. To neglect this possibility does seem to me to constitute a form of dogmatism in Kant's sense, a refusal of the kind of critique that has made modern philosophy ... what it is' (*Essays and Reviews*, 385).

²⁵² From A. W. Moore's commentary on the text in Williams, *Ethics*, 216.

²⁵³ For a brief and illuminating discussion of the epigraph's significance, see Paul Seabright, "Character," *London Review of Books*, September 5, 1985. In its original context in Camus's *The Fall*, the remark appears as an ironical aside referring to the methodical diligence that made the Holocaust possible.

not already living a life and is not already possessed of a set of dispositions sufficiently deep to deserve the name ‘character’. In the first instance, it will be these dispositions that are expressed in ethical reflection: ‘I think about ethical and other goods *from* an ethical point of view that I have already acquired and that is part of what I am.’²⁵⁴

Williams of course rejected the traditional alternative to this way of thinking about what is expressed in ethical reflection. The old notion of a special, and God-given, ‘faculty’ of intuition that allowed non-inferential access to, indeed knowledge of, basic ethical truths did not survive the 19th-century attacks on it by Mill; the notion that Sidgwick would later defend had little in common with the theistic conception.²⁵⁵ If Williams added something to that critique, it was in insisting on a properly historical account of how we came to have the beliefs that some thought a deliverance of Intuition and others a fact about their ‘self-evidence’. In this, he was equally critical of the use to which Sidgwick and his successors (notably Rawls) put the notion of an intuition. He did however declare himself sympathetic to a view he labelled ‘methodological intuitionism’, and took it to be a sufficient condition for an

²⁵⁴ Williams, *Ethics*, 51. For a helpful discussion of why this difference is a more serious

²⁵⁵ The best account of this period of intellectual history is J. B. Schneewind, *Sidgwick’s Ethics and Victorian Moral Philosophy* (Oxford: Oxford University Press, 1977), esp 63–88; and Jerome B. Schneewind, “Whewell’s Ethics,” *Studies in Moral Philosophy, American Philosophical Quarterly Monograph Series* 1 (1968): 108–41.

ethical view to be intuitionist in the relevant sense ‘if it admits a plurality of first principles that may conflict, and, moreover, it has no explicit method or priority rules for resolving such conflicts’.²⁵⁶ Williams thought such a view ‘correctly describes our experience’, but that left open the possibility that our experience ought not to be left in that state – that is to say, we ought to be critical of our intuitions. But what form ought this critique to take?

[I]t is a mistake to think that in order to take a critical view of our ethical beliefs, we have to systematize them in a theoretical style. Indeed, the theoretical stance may not even encourage a critical view. Some systematizations are themselves very conservative, and the rhetoric of radical rationality conceals how conservative they are. The picture of rationality for ethics expressed in terms of theory and system is inadequate. Moreover, we do not have to think that what is principally wrong with our ethical life and our understanding of it is that they are insufficiently rational: they may be, for instance, insufficiently honest.²⁵⁷

These will by now seem familiar thoughts, but it is worth calling attention to William’s qualifications. ‘Some systematizations’, he says, calling to mind his claim quoted earlier that ‘a powerful philosophical theory *can* be an effectively articulated expression of ... distortions’ in our ethical experience, thus ‘blocking radical reconsideration of our morality’.²⁵⁸ But as Martha Nussbaum wrote in her review of *Making Sense of Humanity*, ‘Of course it *can*. But why should we think this is inevitable?’ Why may ‘good theory’ not be used to ‘drive out bad theory’? She points to the uses to which feminist theory (as exemplified in, e.g., the works of Catherine

²⁵⁶ Williams, *Making Sense*, 182.

²⁵⁷ Williams, *Making Sense*, 183.

²⁵⁸ Altham and Harrison eds., *World, Mind, Ethics*, 218–9

MacKinnon) and the liberalism in Mill and Kant, have been used in radical political movements, and concludes that the ‘bold simplicity of theory contains many dangers, but it also contains the potential to organize and transform perceptions on a large scale, in a way that may sometimes be crucial for political change and individual self-criticism.’²⁵⁹

To this charge, it would have been open to Williams simply to turn Nussbaum’s charge around: of course theory *might* do good, but why think *that* inevitable, given that its political record is (for all the putative ethical triumphs of the last century) so decidedly mixed.²⁶⁰ But a better response, consistent with his earlier pronouncements on the question, would be for him to accept that there might be such a good theory, only it would be a theory of rather different sort from that pursued in Anglophone moral philosophy. (Indeed, Catherine MacKinnon was not concerned to answer questions about the criteria of right action in general.) The

²⁵⁹ Martha Nussbaum, “Review of Making Sense of Humanity and Other Philosophical Essays, 1982-1993. by Bernard Williams,” *Ethics* 107, no. 3 (April 1, 1997): 529. These claims are elaborated in Martha Nussbaum, “Why Practice Needs Ethical Theory: Particularism, Principle, and Bad Behaviour,” in *Moral Particularism*, ed. Brad Hooker and Margaret Little (Oxford: Oxford University Press, 2000).

²⁶⁰ As he puts it in relation to utilitarianism as a political theory (Williams, “Critique”, 148): ‘to exercise utilitarian methods on things which at least seem to respond to them is not merely to provide a benefit in some areas which one cannot provide at all. It is, at least very often to provide those things with prestige, to give them an unjustifiably large role in the decision, and to dismiss to a greater distance those things which do not respond to the same methods.’ For a wide-ranging examination of Williams’s views on what kinds of moral considerations might have force in politics, see Edward Hall, “Contingency, Confidence, and Liberalism in the Political Thought of Bernard Williams:,” ed. Margaret Dancy, Victoria Costa, and Joshua Gert, *Social Theory and Practice* 40, no. 4 (2014): 545-69.

pro-theoretical thought was present in his critiques of ordinary language philosophy on the grounds that the mere assembly of linguistic data was useless without theoretical considerations informing the process, and it appears in the preface to *Moral Luck*, where he says, ‘Moral philosophy certainly needs the benefits of theory, but of theory in other parts of philosophy. I am more than ever convinced that what it does not need is a theory of its own.’²⁶¹ If an ethical theory is taken to be ‘a philosophical structure which, together with some degree of empirical fact, will yield a decision procedure for moral reasoning’,²⁶² then no such theory is needed. But he was happy to say even in that Preface that his ideas as they then were stood in need of ‘some rather more systematic framework ... I do not think ... that such a framework could have helpfully preceded these ideas – if there is anything in them, then they have to shape it.’²⁶³

The more extensive position Williams laid out in *Ethics and the Limits of Philosophy* was only the beginning of the process of situating his various local critiques into a more ambitious picture of (moral) philosophy. The most systematic constructive work Williams wrote was the last to be published in his lifetime, *Truth and Truthfulness*, and its contents, unhappily, were abridged and shaped by his

²⁶¹ Williams, *Moral Luck*, ix.

²⁶² Williams, *Moral Luck*, x.

²⁶³ Williams, *Moral Luck*, ix.

knowledge of the seriousness of his illness. But that the work is systematic can hardly be in doubt: it opens with an account of trends in contemporary thought and politics that make an inquiry into the question of truth so pressing, the insistence on debunking and exposure on the one hand and the (post-modern, relativist) scepticism about truth on the other. It offers a statement of its aim, to show that truth, and the dispositions of truthfulness associated with it, have a value that can be revealed by showing them to further basic human interests and needs that can themselves be usefully represented by means of a 'fictional genealogy'. It continues from its imaginary, state-of-nature, story, into real – that is to say, non-fictional – historical narrative that shows how certain episodes in western history gave to the basic dispositions of truthfulness a content that could not be explained or justified simply by the conditions that obtained in the conditions of the fictional genealogy.

Most notably in its discussion of the value of truth-telling, it gives a systematic account of how to make sense of our attitudes to telling and expecting the truth without falling prey to the many temptations to distort those attitudes, either in the Kantian direction of (to take the most notorious example) claiming we have a duty to tell the axe-murderer where to find his victim, or in the (act-)utilitarian direction of thinking there are no general truths on the question at all. Williams in these passages offers a rich, and constructive, account of the many ideas relevant to

theorising a general disposition to sincerity in utterance – notions of trustworthiness, reliability, desert, honour, self-respect, regret, shame, and so forth. Nothing in these passages can be boiled down to a claim of the form ‘S ought to tell the truth iff...’ – but everything up to this point in Williams’s corpus has been suggesting that any way of completing that sentence will be either trivial or inadequate. Near the conclusion of his discussion of the disposition of Sincerity (the capital S marks its use as a term of art), Williams writes:

Sincerity is a disposition, and it cannot be understood just as a disposition to follow a rule. Of course, there have to be some general considerations to which Sincerity attends, or the disposition would have no content. We have just considered some of them. But they do not add up to a rule, in the traditional sense of a requirement which is relatively simple and does not leave most of the work to be done by judgement. If there were a rule, what would it be?²⁶⁴

The penultimate chapter of the book offers the statement of a ‘critical theory principle’, a means by which social practices and institutions may be subjected to critique in part by pressing hard questions about whether they can be regarded as legitimate by those who understand the truth about their origins. What he gave us in *Truth and Truthfulness* was, I submit, a *critical* theory.

²⁶⁴ Williams, *Truth and Truthfulness*, 120–21. The account of the virtue of Sincerity and the broader project of this book seem to me to have all the elements that Nussbaum, “Ethical Theory”, 232–6 requires of an ethical theory: recommendations about practical problems, tests for the correctness of principles, the systematisation and extension of beliefs, some degree of abstractness and generality, potentially universalisable, and a degree of explicitness.

4. Impressionistic Ethics

An interviewer's query once provoked Williams into proposing a unusual model for moral philosophy:

Impressionism was called the painting of modern life ... It didn't look as if it was made out of the old salon material, it looked like something people were actually doing ... I would like there to be moral philosophy that was a bit like Seurat. Something that was directly related to everyday life, but that also made it look rather strange and new because it was ... monumental and had a very strong structure ...²⁶⁵

What could it mean for there to be a moral philosophy 'a bit like Seurat', that pioneer of pointillism and artistic post-Impressionism? What precisely is the contrast here between the artistic *ancien régime* and its 'salon material' on the one hand and the painting of what 'people were actually doing' on the other? What would it be for moral philosophy to be 'monumental' and have 'a very strong structure' even while it maintained its relations to 'everyday life'? And what could be said for such a moral philosophy? What would be – if one is forgiven the pun – the *point* of it?

The point is 'to sharpen perception', we have heard him say, a remark that helpfully underlines the likely motivation for his analogy with the visual arts. In fact, the emphasis on perception can found in a remark in *Ethics and the Limits of Philosophy*, this time directed to the phenomenology of moral experience:

²⁶⁵ Davies, "Fugitive".

There could be a way of doing moral philosophy that started from the ways in which we experience our ethical life. Such a philosophy would reflect on what we believe, feel, take for granted; the ways in which we confront obligations and recognize responsibility; the sentiments of guilt and shame. It would involve a phenomenology of the ethical life. This could be a good philosophy, but it would be unlikely to yield an ethical theory.²⁶⁶

The reason it would be unlikely to do so, the passage continues to say, comes from the nature of ethical theory understood, as it is commonly understood by those who take it to be the point of ethics, as a structure of propositions. Much of this is incipient in his early statement in the preface to *Morality* of the theoretical foundations of his style, that the philosopher's 'initial responsibilities should be [*inter alia*] to moral *phenomena*, as grasped in one's own experience and imagination'.²⁶⁷ Given this pronouncement, it was at least puzzling – as Alasdair MacIntyre was among the few to point out, in a celebratory review of *Moral Luck* – that 'he nowhere draws upon and rarely alludes to either Hegelian or phenomenological attempts to characterize such experience'. MacIntyre hinted at a possible explanation for this reluctance: 'Because Williams shares the idiom, and the standards of clarity, rigor, and relevance of analytic philosophy, it is very plain when he scores a palpable hit. As he often does'²⁶⁸. This suggests that it was not any aversion to the tradition of continental phenomenology that explains his sullen

²⁶⁶ Williams, *Ethics*, 93.

²⁶⁷ Williams, *Morality*, 13; emphasis added.

²⁶⁸ MacIntyre, "Magic", 113.

loyalty to the analytic philosophy in which he was educated, but a sense of the pleasures and virtues of his own tradition.

Williams did take an interest in the ‘continental’ tradition,²⁶⁹ and wrote, albeit seldom and never at length, about Hegel, Sartre and Foucault. His published works show no sign of engagement with Schopenhauer, Husserl or Merleau-Ponty and he seems not have thought much of Heidegger.²⁷⁰ On the other hand, Nietzsche was the one figure for whom Williams came nearest to unequivocal admiration. It is the Nietzschean influence that gives to Williams’s prose the Dionysiac edge that distinguishes it from the purely Apollonian graces of his colleagues. To some degree then, certainly, Williams’s commitment to the analytic style was an accident of his circumstances. A different undergraduate training – if one can ever speculate intelligently on these matters – might have placed his stylistic centre of gravity in a different place.²⁷¹ Still, when he writes in the preface to *Moral Luck* quoted earlier of

²⁶⁹ Of course, he famously disavowed the traditional mode of marking the distinction as involving ‘a cross-classification between the methodological and the geographical: it is like classifying cars as Japanese and front-wheel drive’ (*Philosophy as a Humanistic Discipline*, p. 201).

²⁷⁰ Heidegger was, Williams wrote in 1981, ‘the only world-famous philosopher of the 20th century about whom it can seriously be argued that he was a charlatan’ (Williams, *Essays and Reviews*, 183); I think the remark should be read literally: he is not saying that Heidegger was a charlatan, only that it could be seriously argued that he was, an agnostic claim.

²⁷¹ See also Raymond Geuss, “Did Williams Do Ethics?,” *Arion* 19.3, no. Winter (2012): 149–51 for speculations about which Williams’s serious interest in Nietzsche did not bring him closer to the thinkers of the so-called ‘Frankfurt School’. Much in Geuss’s description of Williams is simply belied by the evidence of the published work, but he is probably right to observe that the figures of the Frankfurt School were among those ‘Continental’ philosophers ‘not sufficiently close to him in temperament and traditional intellectual formation for him to have interacted with them in a

how his scepticism about moral theory had led him ‘to try to find out – often by the crude method of prodding it – which parts of moral thought seemed to me to be actually alive’,²⁷² one is naturally reminded of Nietzsche’s hammer, or tuning fork.²⁷³

It is difficult to avoid taking these *obiter dicta* on method, once juxtaposed, as continuous. It is at least worth asking whether and how they may, taken together, be found to express the different elements of a single vision. The 1985 passage, with its resistance to taking the physical sciences as an exemplar for ethics, helps us make better sense of the 1996 remark to the interviewer, with its hope for an ethics modelled instead on the visual arts. The metaphors of seeing, listening, and prodding give us a sense – but so far no more than a sense – of what the practice of such a moral philosophy would involve, but they do suggest an affinity for at least that strand in ‘Continental’ philosophy that takes philosophy to benefit from drawing upon literature and visual art.

philosophically productive way’ (150). For a forceful argument against Geuss’s reading of Williams, see Paul Sagar, “A Broken Clock,” *The Oxonian Review* 25, no. 3 (June 3, 2014). For some rare, and brief, discussion of thinkers of the Frankfurt School in Williams, see *Truth and Truthfulness*, 3–4 and 304–5 n. 9. This is yet another of the comparisons that must fall (like those with Murdoch and Cavell) outside the scope of this essay.

²⁷² Williams, *Moral Luck*, x.

²⁷³ Williams, *Sense of the Past*, 329 deals with this Nietzschean metaphor: ‘when [Nietzsche] said ... that he philosophized with a hammer, he meant ... that he was testing the wheels, not building a new vehicle. It is not just a question of whether we can foresee a world beyond ours. There is also a question of why we should recognize it as an improvement, as a higher alternative to the life we have.’

The earliest clear example of this practice is to be found in the stories of ‘George’ and ‘Jim’ that remain the most influential, because most memorable, part of his ‘A Critique of Utilitarianism’ – first published as a companion piece to J. J. C. Smart’s ‘An Outline of a System of Utilitarian Ethics’.²⁷⁴ I will come to the stories in a moment, but it is worth reminding ourselves of the prefatory reflections that precede, and explain the point of, the two stories:

... utilitarianism, properly understood and consistently carried through, is a *distinctive* way of looking at human action and morality. ... Where I have offered examples ..., the aim is not just to offer or elicit moral intuitions against which utilitarianism can be tested. ... the aim of the examples and their discussion is not just to ask a question about [what implications one would be willing to live with] and wait for the answer: rather, the aim is to lead into reflections which might show up in greater depth what would be involved in living with these ideas. The first question for philosophy is not “do you agree with utilitarianism’s answer?” but “do you really accept utilitarianism’s way of looking at the question?”²⁷⁵

Note the light-footed move, in the first of the sentences quoted, to considering the utilitarian ‘system of ethics’ Smart has offered as a ‘*way of looking*’; the phrase recurs in the last sentence quoted in Williams’s account of what he takes to be the important question about utilitarianism. The examples follow, and both are worth quoting verbatim and in their entirety; they lose much, more than is usually noticed, in paraphrase. The first example invites us to consider the situation of George the chemist.

²⁷⁴ It is worth noting the word ‘critique’ in Williams’s title; its being published as the ‘Against’ section of a book called *Utilitarianism: For and Against* has drawn attention away from the many other senses in which it is a critique of utilitarianism, i.e. a sustained examination of utilitarianism’s most basic presuppositions. The considerations I have assembled in the previous section give us some reason to entertain the thought that the word was not chosen innocently of these connotations.

²⁷⁵ Williams, “Critique”, 78.

George, who has just taken his Ph.D. in chemistry, finds it extremely difficult to get a job. He is not very robust in health, which cuts down the number of jobs he might be able to do satisfactorily. His wife has to go out to work to keep them, which itself causes a great deal of strain since they have small children and there are severe problems about looking after them. The results of all this, especially on the children, are damaging. An older chemist, who knows about this situation, says that he can get George a decently paid job in a certain laboratory, which pursues research into chemical and biological warfare. George says that he cannot accept this, since he is opposed to chemical and biological warfare. The older man replies that he is not too keen on it himself, come to that, but after all George's refusal is not going to make the job or the laboratory go away; what is more, he happens to know that if George refuses the job, it will certainly go to a contemporary of George's who is not inhibited by any such scruples and is likely if appointed to push along the research with greater zeal than George would. Indeed, it is not merely concern for George and his family, but (to speak frankly and in confidence) some alarm about this other man's excess of zeal, which has led the older man to offer to use his influence to get George the job ... [ellipsis in original] George's wife, to whom he is deeply attached, has views (the details of which need not concern us) from which it follows that at least there is nothing particularly wrong with research into CBW. What should he do?²⁷⁶

The second, and better-known, example concerns Jim the botanist:

Jim finds himself in the central square of a small South American town. Tied up against the wall are a row of twenty Indians, most terrified, a few defiant, in front of them several armed men in uniform. A heavy man in a sweat-stained khaki shirt turns out to be the captain in charge and, after a good deal of questioning of Jim which establishes that he got there by accident while on a botanical expedition, explains that the Indians are a random group of the inhabitants who, after recent acts of protest against the government, are just about to be killed to remind other possible protestors of the advantages of not protesting. However, since Jim is an honoured visitor from another land, the captain is happy to offer him a guest's privilege of killing one of the Indians himself. If Jim accepts, then as a special mark of the occasion, the other Indians will be let off. Of course, if Jim refuses, then there is no special occasion, and Pedro here will do what he was about to do when Jim arrived, and kill them all. Jim, with some desperate recollection of schoolboy fiction, wonders whether if he got hold of a gun, he could hold the captain, Pedro and the rest of the soldiers to threat, but it is quite clear from the set-up that nothing of that kind is going to work: any attempt at that sort of thing will mean that all the Indians will be killed, and himself. The men against the wall, and the other villagers, understand the situation, and are obviously begging him to accept. What should he do?²⁷⁷

Both examples together offer grounds for the now well-worn charges against utilitarianism: that it fails to acknowledge the significance of the difference

²⁷⁶ Williams, 'Critique', 97–8.

²⁷⁷ Williams, 'Critique', 98–9.

between doing something and acting so that someone else does it, that it cannot make sense of 'integrity' as a value, and that it alienates the agent from the actions and the projects that give her life meaning. Before we get into these aspects of the examples, we should pause for a moment on the style in which they are presented. Consider, to start with, the fact that they both involve named protagonists. Jim and George belong to a different literary genre than (say) Russell's omnipresent 'Mr. Jones'. We are, quite literally, put straightaway on first-name terms with them. We should not make too much of this fact; it might reflect no more a generational shift in British social norms. But consider also what else Williams tells us, or implies, about them. Their names mark them as male; there is nothing to suggest they are Caucasian but it is hard to imagine them being anything else; their status as (what are now called) early-career academics puts them squarely, if a little insecurely in George's case, into the middle class; and they are young, likely in their early 30s (George has just received his PhD and has young children; Jim is fit enough to be on a physically arduous expedition). They are also, both, scientists – and therefore, presumably intelligent without being moral philosophers (which would rather change our view of the terms they might be presumed to employ in their deliberation). Again, these facts might not be significant, but I suspect they are more significant than they are usually taken to be. Consider also the (even more conspicuously than is usual in philosophy) *in medias res* opening of Jim's story ('Jim

finds himself...'), the wording full of gratuitous details – the terror and defiance of the Indians, the sweat stains on the captain's khaki uniform – details on which no conclusion turns. Is there a reason for their being there?

I submit this suggestion: The gratuitous details are there to remind – potentially – the reader that any real incident with this sort of structure would be characterised by details of this sort. People have lives and personalities and habits and concerns that they bring with them to the business of making ethical decisions; it is never a pure, deracinated will (as Iris Murdoch might have put it) that makes these decisions. Further, their (i.e. the details') being included, when they need not have been, reminds one that other details have been *excluded*, and therefore, that the very business of inclusion and exclusion already betrays an ethical outlook.

George, in particular, is in a position comparable to that of many of Williams's readers – students, often graduate students, of philosophy, who will shortly have to make life-shaping decisions about their careers. It is significant, also, how much of his narrative involves the use of the free indirect style. There is no quoted dialogue, but the voice is clearly that of the concerned older chemist, and Williams repeatedly draws attention to that voice ('to speak frankly and in confidence' – one can hear him dropping his voice at this point). Williams describes a situation of *advice*; we

know what the older chemist thinks George should do. But what do we – Williams’s readers – think he should do? As Williams would later put it in a reply to the contributors to a 1995 *festschrift*, the point of the question he had asked, ‘ingenuously’, about George and Jim – ‘what ought they to do?’ – was to ask the reader to ask themselves that question.

This is basically, if not perhaps undistractedly, what I had in mind in offering those stories and the discussion of them. ... I brought into the discussion a particular notion, integrity, which I supposed, as others have supposed, might apply in such situations, and I claimed that the application of that notion would stand in the way of the Utilitarian argument and might well stand in the way of the Utilitarian conclusion. I suggested that for someone who used this notion, Utilitarianism’s conclusion might seem downright unacceptable in George’s case, though not so, perhaps, in the other. Over all this, or round it, and certainly at the end of it, there should have been heard ‘what do you think?’, ‘does it seem like that to you?’, ‘what if anything do you want to do with the notion of integrity?’ ... those questions have not been heard. If they have not been heard, or not heard in the right way, that is in good part my fault, because I did not see steadily enough what I wanted the discussion to do.²⁷⁸

A late lecture (‘What might philosophy become?’) has Williams return to some familiar ideas when he says:

Philosophy’s aims ... require that the work sound right, and so they demand an attention to one’s own words. ... Yet this attention has often been one-sided – one-eared, perhaps one might say. We encourage, rightly, a concern with whether it is true ..., but less with whether it *rings* true. A good question, at least to start form, is whether what one has written is something that a grown-up, concerned, intelligent person might say to another about these subjects. Of course it is not the only question ... Philosophy is ... the extension of our most serious concerns by other means, but at least it should introduce our ordinary concerns in a humanly recognizable form. Of much philosophy purportedly about ethical or political subjects ... one may reasonably ask: what if someone speaking to me actually sounded like that?²⁷⁹

²⁷⁸ Altham and Harrison eds., *World, Mind, Ethics*, 210–11.

²⁷⁹ Williams, *Humanistic Discipline*, 206.

By inviting us to ask what George might do just after having heard one concerned grown-up offer him a piece of advice, Williams tries to simulate a parallel situation of responsibility. What advice could we, seriously and in good faith, give him?

We find a different use of the free indirect style in the story of Jim, and we can get some hold on what Williams was trying to do with it if we consider the trenchant remarks of Roy Holland who, no utilitarian, found in the story of Jim everything that was objectionable about the utilitarian way of thinking about ethics. Quoting the captain's words to Jim – the example is in reported speech but the words are clearly those of the captain – he asked,

What is the meaning of this bland tone? Who is this captain? What is this alleged *honouring* of the visitor with the *happy* idea of such a *privilege* as is spoken of? I do not think we should take the dressing in our stride so to speak for the sake of the example. Rather we should consider the role it plays, which is that of providing a 'plausible' background – though I have to put the word 'plausible' in quotes because I do not find much plausibility in the example myself – against which a man who is not supposed to be mad or devilish might say, might utter the threat: 'either you kill this one man or I shall kill twenty'. In fact the idea of the 'honoured guest's privilege' takes the place of what we call at home the motive for the blackmail; but it does so in a way that makes the association with blackmail seem irrelevant, and in this way it provides the framework from within which the visitor's predicament is represented.²⁸⁰

Holland's point, echoed in many respects by Raimond Gaita,²⁸¹ raises an intriguing and immensely serious question, but the passive construction of the last line of the quotation above ('the framework from within which the visitor's predicament is

²⁸⁰ R. F. Holland, "Absolute Ethics, Mathematics and the Impossibility of Politics," *Royal Institute of Philosophy Lectures* 11 (1977), 185–6; emphases in original.

²⁸¹ Raimond Gaita, *Good and Evil: An Absolute Conception* (Routledge, 2004), 64–73.

represented') is ambiguous. Represented by whom? By the captain or by Williams? If by the captain, then it is clear why such a presentation would occlude the fact about the captain's 'offer' being a piece of blackmail: those are not the terms in which the captain sees it. Or more likely, they are not how he chooses to present the choice. This is itself a symptom either of a diseased individual psychology – which Williams does not rule out – or an intelligible portrayal of what the possession of unaccountable power can do to one. If it is Williams who is supposed to be presenting the case within such a framework, that seems to me a misreading of his use of the free indirect style. It is far from clear that Williams was endorsing this way of framing the choice; it just happens that these are the terms in which the captain presents it to Jim. The alleged blandness of the tone in which the example is set out, a consequence of the reported speech in which all the spoken dialogue is couched, was, I think, supposed to be ironic.

There is something to be said for the view that irony is out of place here – on any account, what we have on our hands is a tragedy and a paradigm of political evil. But evil, a notion with deep roots in theology, commands a respect from Gaita and Holland (and indeed Anscombe, who wrote some of her most quotable prose

censuring philosophy that had no room for it²⁸²) that Williams could not have given it. It is true that Williams sometimes acknowledges the human seriousness of the situation – as when he says, ‘If the captain had said on Jim’s refusal, “you leave me with no alternative”, he would have been lying, like most who use that phrase’,²⁸³ or ‘a grounded decision, in Jim’s case ... might not ... be [possible or] even ... decent’.²⁸⁴ But there remains something comic-bookish, or to use the literary genre Jim is briefly aware of inhabiting, Boys’ Own-ish, about the whole thing.

A case like Jim’s could have been rendered in less jaunty, more solemn, tones, and thereby served to engage our sympathies – as even philosophers’ schematic renderings of the horrifying climax of William Styron’s *Sophie’s Choice* (1979) can manage to do. But the fact that it is rendered instead in the way that it is suggests that Williams might not have intended to engage our sympathies at all. The generic tropes that Jim’s story retains from its Boys’-Own antecedents mean that there is a limit to how solemn we can be about it. In this, it is George’s more workaday, more

²⁸² Gaita, *Good and Evil*, 72: ‘the most important philosophical question concerning such examples is not, ‘What ought to be done?’ The most important question is how to characterise the situation and to capture the evil in it. Consequentialists take the evil out of it.’ For Anscombe’s views on the absence of ‘evil’ as a category in twentieth-century moral philosophy, see G. E. M. Anscombe, “Does Oxford Moral Philosophy Corrupt Youth?,” in *Human Life, Action and Ethics: Essays by G. E. M. Anscombe*, ed. Mary Geach and Luke Gormally (Exeter: Imprint Academic, 2006), 161–68; and Anscombe, “Modern moral philosophy”, 17 n. 1.

²⁸³ Williams, “Critique”, 109.

²⁸⁴ Williams, “Critique”, 118.

‘relatable’ dilemma that is successful. (Neither Gaita nor Holland address George’s case directly.) Williams comes close to addressing this sort of worry when he says:

The problem [of finding a style for moral philosophy] ... is to reconcile the general or conceptual concerns of philosophy with a sense of concrete detail, a sense that must be there [if] writing about ethics is going to be realistic or interesting. ... Often, some theory has been under criticism, and the more particular material has come in to remind one of the unreality and, worse, distorting quality of the theory. The material itself, by the standards of good biography, some history, or certain kinds of nineteenth-century fiction, is itself extremely schematic, but it is at least nearer to psychological and social reality than the theory, and helps to bring out the basic point that, very often, the theory is frivolous, in not allowing for anyone’s experience, including the author’s own. Alternatively, the theory does represent experience, but an impoverished experience, which it holds up as the rational norm – that is to say, the theory is stupid.²⁸⁵

One need hardly say that Williams did not think the boys’ adventure story the best model for bringing psychological realism to ethics. I think his point was more subversive: it was that *even* so signally unsophisticated a literary genre was ‘nearer to psychological and social reality’ than some ethical theory, and could be profitably put to use exposing, and helping to correct, its distortions. In this, Williams seems to have followed the practice, if not the theory, of J. L. Austin: a more varied diet of examples would be best, but bad philosophy is guilty of failing to describe adequately even the equivalents of boiled potatoes.²⁸⁶

²⁸⁵ Altham and Harrison eds., *World, Mind, Ethics*, 216–7. For a helpful discussion of the role that fictions of different sorts play in Williams’s political philosophy, see Katrina Forrester, “Judith Shklar, Bernard Williams and Political Realism,” *European Journal of Political Theory*, January 25, 2012, 12.

²⁸⁶ For a classic discussion of this theme that Williams might well have had in mind while writing “A Critique of Utilitarianism”, see Michael Tanner, “Examples in Moral Philosophy,” *Proceedings of the Aristotelian Society*, New Series, 65 (January 1, 1964): 61–76.

Holland is certainly mistaken in attributing to Williams the view that ‘what is central to people’s resistance to consequentialist morality is a concern for personal integrity’,²⁸⁷ as the clarifying remarks quoted above prove. (Williams could doubtless have made it clearer that he did not think integrity, *a fortiori* a concern for *one’s own* integrity,²⁸⁸ essential to the anti-utilitarian case.) The point is rather that some thoughts involving the notion of integrity are one among many things that might keep one from embracing the utilitarian outlook on such a question; the notion of political evil could well be another, though Williams would have been more guarded, as we have seen, about invoking a theological category in this connexion. This is a substantive disagreement about what is important in ethics (one that, I suspect, will roughly track attitudes to Nietzsche) that consequentialism leaves out, and there is something to be said for Williams’s attitudes on the

²⁸⁷ Holland, “Absolute Ethics”, 187. A comparable resistance to Williams’s use of the notion appears in R. M. Hare, “Ethical Theory and Utilitarianism,” in *Utilitarianism and Beyond*, ed. Bernard Williams and Amartya Sen (Cambridge: Cambridge University Press, 1982), 29: ‘Christian humility and *agape* and their humanist counterparts are, then, according to Williams’s standards, an absurd demand ... What is more remarkable is the boldness of the persuasive definition by which he labels the self-centred pursuit of one’s own projects “integrity” and accounts it a fault in utilitarianism that it could conflict with this.’

²⁸⁸ See Williams, *Moral Luck*, 49 for an explicit rejection of this view (‘if [integrity] is regarded as a motive, it is hard to reconstruct its representation in thought except in the objectionable reflexive way: the thought would have to be about oneself and one’s own character, and of the suspect kind. If integrity had to be provided with a characteristic thought, there would be nothing for the thought to be about except oneself – but there is no such characteristic thought, only the thoughts associated with the projects, in carrying out which a man may display his integrity.’ See also Williams, *Ethics*, 10–11, and the illuminating discussion of Williams’s claims on the reflexivity of integrity in Moran, *Authority and Estrangement*, 167 ff.

question.²⁸⁹ Jim and George are a way to ‘prod’ the notion of integrity to see if it has any life in it. Those convinced by Williams will agree with him that it does, but it remains a serious question just how much life it has, and what use it can be put to now.

Stephen Mulhall raises a different concern about the stories of George and Jim, asking where it falls among the two categories Cora Diamond has usefully demarcated. Are they ‘exploration problems’ or are they ‘well-posed problems’? (Roughly, the latter have, given enough information, a unique solution; the former either do not have one, or do not depend for their usefulness on having one.²⁹⁰) Mulhall argues that there is a tension between the exploratory aims of Williams’s stories and the need for them to be, at least from a utilitarian perspective, ‘well-posed’. Mulhall writes:

... the basic purpose of Williams’s tales in fact imposes incoherent demands on their construction. On the one hand, his two imaginary situations must raise questions to which a straightforward application of the utilitarian method will give a unique answer – single, clear, and obvious. On the other hand, those same situations must embody something morally significant that the utilitarian is constrained to overlook. But the first demand requires that Williams construct a well-formed problem from the utilitarian point of view;

²⁸⁹ An intriguing connection: Jim’s situation is roughly paralleled in a scene in David Lean’s *Lawrence of Arabia* (1962) – an amalgam of different episodes in T. E. Lawrence’s *Seven Pillars of Wisdom* – when Lawrence must execute one of his Arab soldiers himself rather than allow him to be killed by a soldier from a rival clan and set off an internecine bloodbath that would irreparably harm the Arab cause. Lawrence does shoot the man and later admits that he found himself enjoying it. It seems to me that Williams is better equipped to make Lawrence’s sentiment intelligible, and to explain why it is not simply a piece of moral depravity.

²⁹⁰ Cora Diamond, “What If X Isn’t the Number of Sheep? Wittgenstein and Thought-Experiments in Ethics,” *Philosophical Papers* 31, no. 3 (2002): 227–50.

whereas the second requires that he construct an exploration-problem – one that allows us to question the assumptions underlying our initial understanding of the problem. What emerges from an exploration of George and Jim’s situation is an ethical value that Williams holds is necessarily excluded from the utilitarian perspective; but then it could not have any place in a tale deliberately constructed so as to form a well-posed problem from that very perspective. In short, integrity is simultaneously absent and present in the two tales, endlessly flickering in and out of our field of vision; and this merely illustrates the impossibility of treating them simultaneously in these two different ways.²⁹¹

Mulhall’s suggestion nicely explains some of the puzzlement the stories have aroused in the commentarial literature they have generated – puzzlement among the non-utilitarians about why Williams incorporates so many utilitarian assumptions into a parable with an ostensibly anti-utilitarian moral, puzzlement among the utilitarians about how Williams proposes to derive an anti-utilitarian moral from stories structured by their assumptions. Certainly, Williams himself seems – as the retrospective *mea culpa* quoted earlier suggests – to have rued much about his presentation of the stories. But there is more to be said on Williams’s behalf on this point, more perhaps than he wished to say himself.

First, both stories are described in ways that allow for redescription in (even) more, or less, utilitarian terms. We hear from George’s friend and are told a little about his wife; we hear from the captain, and we hear from the villagers. The one person we do not hear from in both cases is the person whose choice we are reflecting on: George and Jim. This omission is the vacuum that I take it our reflections are to fill.

²⁹¹ Stephen Mulhall, “The Mortality of the Soul: Bernard Williams’s Character(s),” in *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*, ed. Alice Crary (MIT Press, 2007), 367.

The question is this: what, or whom, are we to be thinking about? It is characteristic of the ‘absolute’ ethics of Gaita and Holland that their thoughts are with the villagers, not with Jim (they say nothing of George, perhaps because it is much less clear whether there are any *particular* individuals who play in his story the part played by the supplicant villagers in Jim’s). However, if we think what Jim or George ought to do depends on some distinctive fact about them, it will not seem precious to suppose that their integrity might be relevant to how we conceive of their situations.²⁹²

Many questions are begged in both men’s stories, as Williams had warned us they would be. To stick with George: what *are* his compunctions, exactly, and what is their source in his past? It would make some difference to how we think of his dilemma whether he was a casual debating-society pacifist or a Quaker and committed demonstrator for the Campaign for Nuclear Disarmament.²⁹³ How much would it cost him to turn down the job? How bad would it be for his wife to have to take a job? Could she come to share his convictions on the matter? How much of a difference can this other candidate’s zeal for chemical and biological warfare really make? How important a contribution does this laboratory make to the activities of

²⁹² See Williams, “Critique”, 104: ‘George’s wife has some reason to be interested in George’s integrity and his sense of it; the Indians, quite properly, have no interest in Jim’s.’

²⁹³ See Williams, “Critique”, 111 for a brief allusion to such questions.

the military-industrial complex? And so forth. To be clear, the narrowness of George's range of choices is not to be blamed on the impoverishment of philosophy's literary resources; it is a feature, and a readily recognisable one, of the choices people have to make all the time.

In fact, I am inclined to think that George and Jim are not best regarded as figures in a *literary* composition,²⁹⁴ but rather as subjects of a painting – a form whose conventions are better suited to do the job Williams wants of them, not to immerse his readers into a fully realised ethical situation but to force an apprehension of certain structural features of the ethical life, features that are only discernible when the scene is rendered in a certain stylised way. George is a realistic figure, but he is being painted in accordance with a special set of conventions. Knowing the facts of human life and modern society to be what they are, it takes no special feat of the imagination to fill in the gaps. This is, in fact, a solution to what Mulhall takes to be

²⁹⁴ It is worth entertaining the thought that the stories of George and Jim are what Williams would, much later in his life, label 'stark' fictions – as exemplified by the tragedies of Sophocles – as opposed to the 'dense' fictions exemplified by the Victorian realist novel. 'It is not merely that [the] style and structure [of a stark fiction] avoid the anecdotal and the incidental, but that these resources are typically directed in a concentrated way to displaying the operations of chance and necessity' (Williams, *Sense of the Past*, 56). George and Jim certainly avoid, though as we have seen rather less so than other such examples in recent moral philosophy, the 'anecdotal and incidental', but what they are directed to dramatising is a situation of *choice*. The question we are asked about George and Jim, 'what should they do?', seems much less *apropos* to the over-determined situations of Sophocles' Ajax or Jocasta. For a helpful set of criticisms of Williams's pessimistic reading of Sophocles, see Martha Nussbaum, "Bernard Williams: Tragedies, Hope, Justice," in *Reading Bernard Williams*, ed. Daniel Callcut (Routledge, 2008); for the claim that Williams's views developed in significant ways between 'A Critique of Utilitarianism' and these late reflections on Greek tragedy, see Mulhall, "The Mortality of the Soul: Bernard Williams's Character(s)", 373–5.

a problem, the way that in the two stories integrity is ‘endlessly flickering in and out of our field of vision; ... this merely illustrates the impossibility of treating them simultaneously in these two different ways’. But why not think this a virtue?

Let us consider Seurat, the artist Williams explicitly alludes to as a potential model for moral philosophers. His most celebrated work, *A Sunday Afternoon on the Island of La Grande Jatte*, exemplifies exactly those qualities – of being monumental, possessing a conspicuous structure, of having the capacity to make the familiar strange – that Williams recommended in a constructive piece of moral philosophy.²⁹⁵ But it might be worth pursuing the analogy with that painting even further in a spirit of experiment, just to see what (if anything) the exercise might reveal. One such element is the crowdedness of the canvas – for all his careful thought to composition, there is no single thing that Seurat forces the viewer to look at. The sheer size of the canvas means that the painting yields different things to the viewer depending on one’s distance from the canvas. Stand right up against it, and the image dissolves into countless tiny (‘pointillist’) brushstrokes; with every step back, more elements are revealed. From a distance, a variety of subjects come into view, a whole scene, richly populated with disparate elements. Being a

²⁹⁵ I am grateful to Patricia Williams for letting me know that Williams did in fact have great affection for this painting to point of insisting on travelling to Chicago’s Art Institute to see it even when he was seriously ill.

painting, Seurat's *Grande Jatte* does not, and cannot, offer any easy window onto the inner lives of those it depicts. But it does not follow that it is uninterested in their inner lives. On the contrary, it makes it part of the experience of viewing the painting that we are invited to conjecture, based on the visible clues the painting does give us, what might be going on in the minds of its subjects. Seurat's painting too has famously raised questions that the image on the canvas itself cannot answer. Take the monkey in Seurat's frame – why is it there? – or the woman fishing in the Seine – is she or is she not soliciting? These are, and must remain, questions. But the value of the questions is something quite different from that of having them answered. It is rather the value of learning to see, and to think, in new ways.

George and Jim are there to enable elements of ordinary, everyday moral thought to stake their claims at the philosopher's table. We might think of them as *details in a diptych* – Jim's story forming the basis for a baroque 'agony in the jungle' scene, George's a more quotidian bourgeois image. We are not simply *given* their inner lives. We are instead invited to wonder about them: what projects and commitments might they have, and how might those projects and commitments figure in their lives, how might those projects impinge on the question of what each ought, in his respective circumstances, to do? We can stand up close and

contemplate the narrow question asked – what ought they to do – or we might step back, and observe the ways in which the two stories illuminate and complicate each other, and then look to the edges of the two scenes for how they point outward.

From Jim's dilemma to the desperation of the villagers, to the jungle from which Jim emerged expecting no more than an evening around the fire and another day collecting specimens, to the evils of a military junta in a faraway capital. From George's dilemma to the reassuring face of the kindly friend, the urgings of his overworked wife, the faces of his miserable children, maybe even the CND membership card he might have stowed away in an old wallet, to the factory itself, and the weapons that go out to military bases far away to be used for goodness knows what purposes. The relations can be put in more abstract terms: both stories are about the ways in which individual choices ramify, outward in the direction of their effects on other people, and inward, in their effects on the self. A number of notions, not just integrity, are 'endlessly flickering in and out of our field of vision' as we consider the images, but it should not be assumed that this is evidence of an irresolvable tension. Once one has stepped back and surveyed the wider scene, our apprehension of the small detail is transformed when we look at it again, up close.

George and Jim do not embody any kind of triumphal celebration of ordinary moral thought and its folksy virtues; their stories are part of an attempt to bring up the

question, too quickly put aside, of whether the discursive constraints imposed by certain kinds of philosophical reflection are fair, and whether something of intellectual and ethical importance has been lost along the way. If George and Jim do anything, it is by bringing such notions as that of integrity, but not only integrity, back into a conversation from which they had been – not banished, exactly, but left unheard, because other notions (duty, utility) were speaking so loudly; or to return to our artistic metaphor, because those other notions were dominating the canvas. Williams saw it, in other words, as part of philosophy's business to give us the terms in which to express what we previously saw vaguely, to make real what was little more than a lingering disquiet at the edge of our consciousness: this is yet another sense in which his critique of utilitarianism was a critique of the Kantian kind, surveying the landscape of moral experience to see what needed to be present there – or absent – in order for the deliverances of utilitarian thought to be justified.

5. Style and Stylishness

Williams was aware of, and to some degree endorsed, a prominent thread in other writers that conceived of the aims of philosophy in decidedly deflationary terms: 'Philosophy ..., on lines variously laid down by Hume, Wittgenstein, Stanley Cavell,

needs to recall us to the everyday', he once wrote.²⁹⁶ In this, he has some claim to being a conservative in his moral philosophy. But there is more to his aims than this – as indeed there was to those of Hume, Wittgenstein and Cavell. His idea that a good philosophical argument might help 'to sharpen perception' points towards a slightly different, more suspicious, more critical, ideal. For Williams, as for Orwell before him, 'To see what is in front of one's nose needs a constant struggle.'²⁹⁷ As the concluding words of one of his reviews had it, 'no inquiry that is going to help us understand ourselves can do without that kind of truthfulness, an acute and wary sense of the ordinary.'²⁹⁸

Yet, there was in him no romanticism about the common sense or perceptiveness of the yeoman, though some of his sardonic commentary on philosophers' errors can encourage a picture of him as a romantic of this sort. A sense of the ordinary is an *achievement* at the best of times, and it is possible that philosophy can sometimes, to some people, be of help in cultivating it.²⁹⁹ The voice of his best aphorisms – 'this construction provides the agent with one thought too many',³⁰⁰ 'the only serious

²⁹⁶ Williams, *Truth and Truthfulness*, 10.

²⁹⁷ George Orwell, "In Front of Your Nose," *The Tribune*, March 22, 1946.

²⁹⁸ Williams, *Essays and Reviews*, 282.

²⁹⁹ See Williams, *Ethics*, 202: 'Some people might even get help from philosophy in living ... a [meaningful] life—but not ... each reflective person, and not from the ground up.'

³⁰⁰ Williams, *Moral Luck*, 18.

enterprise is living'³⁰¹ – is not the voice of a conservative. One does not hear someone concerned 'to reassure the people that their prejudices are true'.³⁰² One does, rather, hear something of a Kantian echo:

Thus is ordinary human reason forced to go outside its sphere and take a step into the field of practical philosophy, not by any need for speculation (which never befalls such reason so long as it is content to be mere sound reason) but on practical grounds themselves ... Thus when ordinary practical reason cultivates itself, there imperceptibly arises in it a dialectic which compels it to seek help in philosophy. ... in this case, just as in the other [i.e. theoretical reason], peace will be found only in a thorough critical examination of our reason.³⁰³

An essay written very early in Williams's career declared the book under review one in which 'notable intelligence is deployed in distorting the facts of human experience to fit an extremely simple view of human nature, science, and philosophy.' To understand why this was so 'really awaits a proper study of the concept of "common sense" and of the peculiar role it plays in some contemporary philosophy.'³⁰⁴ The hero of this common-sense tradition was G. E. Moore boldly slaying the sceptic by claiming to know he had hands. A less strident defender of common sense, particularly in its moral form, was Moore's Cambridge predecessor, Henry Sidgwick, who aimed to show (as Williams summarised it), 'the only justification of Utilitarianism ... is ... dependent upon certain general principles

³⁰¹ Williams, *Ethics*, 117.

³⁰² This formulation of the role of a conservative intellectual is attributed to Roger Scruton, in Daniel Hannan, "The Paradox of the Conservative Intellectual," *News - Telegraph Blogs* (blog), February 23, 2010, <http://blogs.telegraph.co.uk/news/danielhannan/100027134/the-paradox-of-the-conservative-intellectual/>

³⁰³ Immanuel Kant, *Grounding for the Metaphysics of Morals: With On a Supposed Right to Lie because of Philanthropic Concerns*, trans. James Wesley Ellington (Hackett Publishing, 1993), 17.

³⁰⁴ Williams, *Essays and Reviews*, 8.

which are intuitive', that is to say, derivable from the most basic elements of common-sense morality. Williams found much to admire in Sidgwick's *The Methods of Ethics*, yet he found it

a general feature of the book that the account of commonsense morality is very linear, lacking any sense of the possibility of alternative moral traditions or the idea that certain moral outlooks of cultivated contemporary opinion might represent interests less broad than those of society or mankind as a whole. His discussion for instance of the ethics of making a financial profit out of persons who are in a disadvantaged situation is one example where the notion of *commonsense* itself demands to be examined in ways in which Sidgwick never examines it.³⁰⁵

The sort of antipathy to academic moral philosophy Williams carried with him throughout his long career was not the conservative's hostility to a critical examination of current practices; it was an impatience with an examination not nearly thorough or imaginative enough. The obscure essay of Sidgwick's, an appreciation of the poetic corpus of Arthur Hugh Clough, from which I have drawn my first epigraph has Sidgwick describe what is evidently, to him, an exciting new intellectual trend: 'we more and more say and write what we actually do think and feel, and not what we intend to think or should desire to feel'.³⁰⁶ But Sidgwick's own work in ethics was marred, Williams thought, by its failure to do just this. In Williams's vision of ethics, to find out 'what we actually do think and feel' is not a precursor to the critique that is the real business of philosophy. To find this out, to describe it, truthfully, fully, richly, and in a style responsive to its protean, half-

³⁰⁵ Williams, *Sense of the Past*, 283.

³⁰⁶ Sidgwick, *Miscellaneous Essays*, 60; the full quotation may be read at the start of this essay.

formed character: this is part of the critical enterprise. As in other forms of critique, the aim in Williams is to emancipate the reflective individual, 'from distortions or misunderstandings involved in his or her own experience, and the aim of this ... is the Socratic aim of improving that experience'.³⁰⁷ Williams's Socrates has no ambitions to scientific objectivity, no hope of transcending the merely human; he is a humanist, ironic and warm, an emancipatory figure. Perhaps every generation, eventually, gets the Socrates it needs.

³⁰⁷ Altham and Harrison eds., *World, Mind, Ethics*, 218–9.

Bibliography

Altham, James Edward John, and Ross Harrison, eds. *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*. Cambridge: Cambridge University Press, 1995.

Annas, Julia. *An Introduction to Plato's Republic*. New York: Oxford University Press, 1981.

Anscombe, G. E. M. "Does Oxford Moral Philosophy Corrupt Youth?" In *Human Life, Action and Ethics: Essays by G. E. M. Anscombe*, edited by Mary Geach and Luke Gormally, 161–68. Exeter: Imprint Academic, 2006.

———. "Modern Moral Philosophy." *Philosophy* 33, no. 124 (1958): 1–19.

Austin, J. L. "A Plea for Excuses: The Presidential Address." *Proceedings of the Aristotelian Society*, New Series, 57 (January 1, 1956): 1–30.

———. *Sense and Sensibilia*. Edited by G. J. Warnock. 1st edition. Oxford: Oxford University Press, 1962.

Bailly, Jacques A. "What You Say, What You Believe, and What You Mean." *Ancient Philosophy* 19, no. Special (1999): 65–76.

Barnes, Jonathan. "Aristotle and the Methods of Ethics." In *Method and Metaphysics: Essays in Ancient Philosophy I*, 174–94. Oxford: Oxford University Press, 2011.

———. “Like Us, Only Better.” *The Times Literary Supplement*, no. 4699 (April 23, 1993):

3–4.

———. *Method and Metaphysics: Essays in Ancient Philosophy I*. Oxford: Oxford

University Press, 2011.

Barney, Rachel. “Socrates’ Refutation of Thrasymachus.” In *The Blackwell Guide to*

Plato’s Republic, edited by Gerasimos Xenophon Santas. Oxford: Blackwell Publishing,

2006.

Baz, Avner. *When Words Are Called For*. Cambridge, Mass.: Harvard University Press,

2012.

Benjamin, Walter. *Selected Writings: 1913-1926*. Cambridge, Mass.: Harvard University

Press, 1996.

Benson, Hugh H. “The Dissolution of the Problem of the Elenchus’.” *Oxford Studies in*

Ancient Philosophy 13 (1995): 45–112.

Beverluis, John. *Cross-Examining Socrates: A Defense of the Interlocutors in Plato’s Early*

Dialogues. Cambridge: Cambridge University Press, 2000.

Bobonich, Christopher. *Plato’s Utopia Recast: His Later Ethics and Politics*. Oxford:

Clarendon Press, 2004.

Brandom, Robert. *Making It Explicit: Reasoning, Representing, and Discursive*

Commitment. Cambridge, Mass.: Harvard University Press, 1998.

- Brickhouse, Thomas C., and Nicholas D. Smith. *Plato's Socrates*. Oxford: Oxford University Press, 1994.
- Broackes, Justin, ed. *Iris Murdoch, Philosopher*. Oxford: Oxford University Press, 2011.
- Burnyeat, M. F. "Examples in Epistemology: Socrates, Theaetetus and G. E. Moore." *Philosophy* 52, no. 202 (October 1977): 381–98.
- . "Socratic Midwifery, Platonic Inspiration." *Bulletin of the Institute of Classical Studies* 24, no. 1 (December 1, 1977): 7–16.
- Burnyeat, Myles, and Plato. *The Theaetetus of Plato*. Translated by M. J. Levett. Indianapolis: Hackett Publishing, 1990.
- Cavell, Stanley. *Must We Mean What We Say?: A Book of Essays*. Cambridge: Cambridge University Press, 2002.
- Chappell, Timothy. "Bernard Williams." *Stanford Encyclopedia of Philosophy*, 2014. <http://plato.stanford.edu/archives/sum2014/entries/williams-bernard/>.
- . "Review: Bernard Williams, Philosophy as a Humanistic Discipline, The Sense of the Past." *Philosophical Investigations* 32, no. 4 (October 2009): 360–71.
- Collini, Stefan. *Absent Minds: Intellectuals in Britain*. Oxford: Oxford University Press, 2006.
- Crisp, Roger. "Modern Moral Philosophy and the Virtues." In *How Should One Live?: Essays on the Virtues*, edited by Roger Crisp. Oxford: Oxford University Press, 1998.
- . *Reasons and the Good*. Oxford: Oxford University Press, 2006.

Davies, John. "A Fugitive from the Pigeonhole." *Times Higher Education*, November 4, 1996.

Diamond, Cora. "What If X Isn't the Number of Sheep? Wittgenstein and Thought-Experiments in Ethics." *Philosophical Papers* 31, no. 3 (2002): 227–50.

Dodds, E. R. *The Greeks and the Irrational*. Oakland: University of California Press, 2004.

Dover, Kenneth James. *Greek Popular Morality in the Time of Plato and Aristotle*.

Indianapolis: Hackett Publishing, 1974.

Dummett, Michael. *Truth and Other Enigmas*. Cambridge, Mass.: Harvard University Press, 1978.

Evans, Gareth. *The Varieties of Reference*. Edited by John McDowell. Oxford: Oxford University Press, 1982.

Fink, Jakob L. *The Development of Dialectic from Plato to Aristotle*. Cambridge: Cambridge University Press, 2012.

Fodor, Jerry A. *The Language of Thought*. Cambridge, Mass.: Harvard University Press, 1975.

Forrester, Katrina. "Judith Shklar, Bernard Williams and Political Realism." *European Journal of Political Theory*, January 25, 2012, 1474885111430605.

Foucault, Michel. *The Politics of Truth*. Edited by Sylvère Lotringer. Translated by Lysa Hochroth and Catherine Porter. Los Angeles: Semiotext, 2007.

- Frede, Michael. "Plato's Arguments and the Dialogue Form." Edited by James C. Klagge and Nicholas D. Smith. *Methods of Interpreting Plato and His Dialogues (Oxford Studies in Ancient Philosophy)* Supplementary volume (1992): 201–20.
- Gaita, Raimond. *Good and Evil: An Absolute Conception*. Abingdon: Routledge, 2004.
- Geach, P. T. "Plato's Euthyphro." *The Monist* 50, no. 3 (1966): 369–82.
- Geertz, Clifford. "Common Sense as a Cultural System." In *Local Knowledge*, 73–93. New York: Basic Books, 1983.
- Gellner, Ernest. *Words and Things: An Examination Of, and an Attack On, Linguistic Philosophy*. Victor Gollancz, 1959.
- Geuss, Raymond. "Did Williams Do Ethics?" *Arion* 19.3, no. Winter (2012): 141–62.
- Gifford, Mark. "Dramatic Dialectic in Republic Book 1." In *Oxford Studies in Ancient Philosophy*, edited by David Sedley, XX:25–106. Oxford: Clarendon Press, 2001.
- Gill, Christopher. "Speaking Up for Plato's Interlocutors. A Discussion of J. Beversluis, Cross-Examining Socrates." *Oxford Studies in Ancient Philosophy* 20 (2001): 297–321.
- Guthrie, W. K. C. *A History of Greek Philosophy: Volume 1, The Earlier Presocratics and the Pythagoreans*. Cambridge: Cambridge University Press, 1979.
- Hall, Edith. *Inventing the Barbarian: Greek Self-Definition Through Tragedy*. Oxford: Oxford University Press, 1989.

Hall, Edward. "Contingency, Confidence, and Liberalism in the Political Thought of Bernard Williams:" *Social Theory and Practice* 40, no. 4 (2014): 545–69.

Hare, Richard Mervyn. *Essays on Philosophical Method*. Oakland: University of California Press, 1972.

———. *The Language of Morals*. Oxford: Oxford University Press, 1952.

Hare, R. M. "Ethical Theory and Utilitarianism." In *Utilitarianism and Beyond*, edited by Bernard Williams and Amartya Sen. Cambridge: Cambridge University Press, 1982.

———. *Plato*. Reprint edition. Oxford: Oxford Paperbacks, 1982.

Heuer, Ulrike, and Gerald Lang, eds. *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams*. New York: Oxford University Press, 2012.

Holland, R. F. "Absolute Ethics, Mathematics and the Impossibility of Politics." *Royal Institute of Philosophy Lectures* 11 (1977): 172–88.

Hurka, Thomas, ed. *Underivative Duty: British Moral Philosophers From Sidgwick to Ewing*. Oxford: Oxford University Press, 2011.

Irwin, Terence. "Common Sense and Socratic Method." In *Method in Ancient Philosophy*, edited by Jyl Gentzler. Oxford: Oxford University Press, 2001.

———. *Plato's Ethics*. New York: Oxford University Press, 1995.

———. *The Development of Ethics: From Socrates to the Reformation*. Oxford: Oxford University Press, 2007.

Irwin, T. H. "Say What You Believe." *Apeiron* 26, no. 3/4 (1993): 1–16.

Jeffries, Stuart. "The Quest for Truth." *The Guardian*, November 30, 2002.

<http://www.guardian.co.uk/books/2002/nov/30/academicexperts.highereducation>

Kant, Immanuel. *Grounding for the Metaphysics of Morals (Third Edition): With On a Supposed Right to Lie because of Philanthropic Concerns*. Translated by James Wesley Ellington. Indianapolis: Hackett Publishing, 1993.

———. *Groundwork for the Metaphysics of Morals*. Translated by Allen Wood. New Haven: Yale University Press, 2002.

Kraut, Richard. *The Cambridge Companion to Plato*. Cambridge: Cambridge University Press, 1992.

Laugier, Sandra. *Why We Need Ordinary Language Philosophy*. Chicago: University of Chicago Press, 2013.

Lazari-Radek, Katarzyna de, and Peter Singer. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press, 2014.

Lear, Jonathan. "The Socratic Method and Psychoanalysis." In *A Companion to Socrates*, edited by Sara Ahbel-Rappe and Rachana Kamtekar, 442–62. Oxford: Blackwell Publishing, 2007.

Lewis, David. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8, no. 1 (January 1979).

Long, A. G. *Conversation and Self-Sufficiency in Plato*. Oxford: Oxford University Press, 2013.

Macfarlane, Robert. *The Wild Places*. London: Granta Books, 2008.

MacIntyre, Alasdair. *After Virtue: A Study in Moral Theory*. Notre Dame: University of Notre Dame Press, 1981.

———. “The Magic in the Pronoun ‘My.’” *Ethics* 94, no. 1 (October 1983): 113–25.

———. “Truthfulness, Lies, and Moral Philosophers: What Can We Learn from Mill and Kant?” *The Tanner Lectures on Human Values* 16 (1995): 307–61.

Magee, Bryan, ed. *Talking Philosophy: Dialogues with Fifteen Leading Philosophers*. Oxford: Oxford University Press, 2001.

McCabe, Mary Margaret. *Plato and His Predecessors: The Dramatisation of Reason*. Cambridge: Cambridge University Press, 2000.

McDowell, John. “Non-Cognitivism and Rule-Following.” In *Wittgenstein: To Follow A Rule*, edited by S. Holtzman and Christopher M. Leich, 141–62. Abingdon: Routledge, 1981.

Mikalson, Jon D. *Honor Thy Gods: Popular Religion in Greek Tragedy*. UNC Press Books, 1992.

Moore, A. W. “Williams, Sir Bernard Arthur Owen (1929–2003).” *Dictionary of National Biography*, January 2007. <http://www.oxforddnb.com/view/article/90066>.

———. *Noble in Reason, Infinite in Faculty: Themes and Variations in Kant's Moral and Religious Philosophy*. Abingdon: Routledge, 2003.

———. *The Evolution of Modern Metaphysics: Making Sense of Things*. Cambridge: Cambridge University Press, 2011.

Moore, G. E. *Philosophical Papers*. London: George Allen & Unwin Publishers, 1959.

Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press, 2001.

———. "Problems of Sincerity." *Proceedings of the Aristotelian Society* 105, no. 3 (June 1, 2005): 341–61.

Morrison, Donald R. *The Cambridge Companion to Socrates*. Cambridge: Cambridge University Press, 2010.

Moss, Jessica. "Plato's Appearance-Assent Account of Belief." *Proceedings of the Aristotelian Society* 114, no. 2pt2 (July 1, 2014): 213–38.

Mulhall, Stephen. "The Mortality of the Soul: Bernard Williams's Character(s)." In *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*, edited by Alice Crary, 355–78. Cambridge, Mass: MIT Press, 2007.

Murdoch, Iris. *Existentialists and Mystics: Writings on Philosophy and Literature*. London: Penguin Books, 1999.

———. *The Sovereignty of Good*. Abingdon: Routledge, 2001.

———. “Vision and Choice in Morality.” *Proceedings of the Aristotelian Society: Dreams and Self-Knowledge* Supplement no. 30 (1956): 32–58.

Nails, Debra. *The People of Plato: A Prosopography of Plato and Other Socratics*.

Indianapolis: Hackett Publishing, 2003.

Nehamas, Alexander. “Eristic, Antilogic, Sophistic, Dialectic: Plato’s Demarcation of Philosophy from Sophistry.” *History of Philosophy Quarterly* 7, no. 1 (January 1, 1990): 3–16.

Nietzsche, Friedrich. “*On the Genealogy of Morality*” and *Other Writings*. Edited by Keith Ansell-Pearson. Translated by Carol Diethe. Cambridge: Cambridge University Press, 2006.

Nussbaum, Martha C. “Bernard Williams : Tragedies, Hope, Justice.” In *Reading Bernard Williams*, edited by Daniel Callcut. Abingdon: Routledge, 2008.

———. “Review of *Making Sense of Humanity and Other Philosophical Essays, 1982-1993* by Bernard Williams.” *Ethics* 107, no. 3 (April 1, 1997): 526–29.

———. “Why Practice Needs Ethical Theory: Particularism, Principle, and Bad Behaviour.” In *Moral Particularism*, edited by Brad Hooker and Margaret Little. Oxford: Oxford University Press, 2000.

———. “Socratic Studies.” *The Journal of Philosophy* 94, no. 1 (January 1, 1997): 27–45.

———. *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. 2 edition. Cambridge: Cambridge University Press, 2001.

- . *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*. Princeton: Princeton University Press, 1996.
- O’Grady, Jane. “Professor Sir Bernard Williams.” *The Guardian*. June 13, 2003.
<http://www.guardian.co.uk/news/2003/jun/13/guardianobituaries.obituaries>.
- Orwell, George. “In Front of Your Nose.” *The Tribune*. March 22, 1946.
- Osborne, Catherine. “Socrates in the Platonic Dialogues.” *Philosophical Investigations* 29, no. 1 (2006): 1–21.
- Palfreyman, David, ed. *The Oxford Tutorial: Thanks, You Taught Me How to Think*. Oxford: The Oxford Centre for Higher Education Policy Studies, 2008.
- Parfit, Derek. *On What Matters, Volume I*. Oxford: Oxford University Press, 2011.
- . *Reasons and Persons*. Oxford: Oxford University Press, 1984.
- Payne, Andrew. “The Division of Goods and Praising Justice for Itself in Republic II.” *Phronesis* 56, no. 1 (January 1, 2011): 58–78.
- Pears, David Francis, ed. *The Nature of Metaphysics*. London: Macmillan, 1957.
- Plato. *Complete Works*. Edited by John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett Publishing, 1997.
- Rawls, John. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press, 1971.
- . *Lectures on the History of Political Philosophy*. Edited by Samuel Richard Freeman. Cambridge, Mass.: Harvard University Press, 2009.
- Roberts, Debbie. “Shapelessness and the Thick.” *Ethics* 121, no. 3 (2011): 489–520.

Robinson, Richard. *Plato's Earlier Dialectic*. Vol. 4. 17. Oxford: Oxford University Press, 1953.

Rosenfeld, Sophia. *Common Sense: A Political History*. Cambridge, Mass: Harvard University Press, 2011.

Ross, W. D. *Foundations of Ethics: The Gifford Lectures Delivered in the University of Aberdeen, 1935-*. Oxford: Oxford University Press, 1939.

Ryle, Gilbert. *Collected Papers: Volume I*. Hutchinson, 1971.

———. *The Concept of Mind*. 1st edition. Chicago: University of Chicago Press, 1949.

———. *The Concept of Mind*. 60th anniversary edition. Abingdon: Routledge, 2009.

Sagar, Paul. "A Broken Clock." *The Oxonian Review* 25, no. 3 (June 3, 2014).

———. "Minding the Gap: Bernard Williams and David Hume on Living an Ethical Life." *Journal of Moral Philosophy* 11, no. 5 (September 24, 2014): 615–38.

Sartre, Jean-Paul. *Essays in Existentialism*. Edited by Wade Baskin. New Jersey: Citadel Press, 1965.

Schneewind, J. B. *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Oxford University Press, 1977.

———. "Whewell's Ethics." *Studies in Moral Philosophy, American Philosophical Quarterly Monograph Series* 1 (1968): 108–41.

Schultz, Bart. *Henry Sidgwick - Eye of the Universe: An Intellectual Biography*. Cambridge: Cambridge University Press, 2004.

- Scott, Dominic. *Plato's Meno*. Cambridge: Cambridge University Press, 2006.
- Scott, Gary Alan, ed. *Does Socrates Have a Method?: Rethinking the Elenchus in Plato's Dialogues and Beyond*. Philadelphia: Penn State Press, 2002.
- Seabright, Paul. "Character." *London Review of Books*, September 5, 1985.
- Seanor, Douglas, and N. Fotion, eds. *Hare and Critics: Essays on Moral Thinking*. Oxford: Clarendon Press, 1988.
- Sedley, David. *The Midwife of Platonism: Text and Subtext in Plato's Theaetetus*. Oxford: Oxford University Press, 2004.
- Sidgwick, Arthur, and Eleanor Mildred Sidgwick. *Henry Sidgwick: A Memoir*. London: Macmillan, 1906.
- Sidgwick, Henry. *Miscellaneous Essays and Addresses*. Edited by Eleanor Mildred Sidgwick and Arthur Sidgwick. London: Macmillan, 1904.
- . *Outlines of the History of Ethics for English Readers*. London: Macmillan, 1886.
- . *The Methods of Ethics*. 1st edition. London: Macmillan, 1874.
- . *The Methods of Ethics*. 7th edition. London: Macmillan, 1907.
- Singer, Marcus G. "Sidgwick and Nineteenth-Century Ethical Thought." In *Essays on Henry Sidgwick*, edited by Bart Schultz, 65–91. Cambridge: Cambridge University Press, 1992.
- Singer, Peter. "Sidgwick and Reflective Equilibrium." *The Monist* 58, no. 3 (1974): 490–517.

- Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 1973.
- Strauss, Darin. *Half a Life*. New York: Random House Trade, 2011.
- Strawson, P. F. *Individuals: An Essay in Descriptive Metaphysics*. Methuen, 1959.
- Tanner, Michael. "Examples in Moral Philosophy." *Proceedings of the Aristotelian Society*, New Series, 65 (January 1, 1964): 61–76.
- Tanney, Julia. "Gilbert Ryle." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy*, 2009. <http://plato.stanford.edu/archives/win2009/entries/ryle/>.
- Tarrant, Dorothy. "Plato's Use of Quotations and Other Illustrative Material." *The Classical Quarterly* 1, no. 1–2 (January 1951): 59.
- Tsouna, Voula. "Socrate et La Connaissance de Soi: Quelques Interprétations." *Philosophie Antique* 1 (2001): 37–64.
- Vlastos, Gregory. *Socrates: Ironist and Moral Philosopher*. Ithaca: Cornell University Press, 1991.
- . "The Socratic Elenchus." *Oxford Studies in Ancient Philosophy* 1 (1983): 27–58.
- Voorhoeve, Alex. *Conversations on Ethics*. Oxford: Oxford University Press, 2009.
- Warnock, G. J. *English Philosophy since 1900*. Oxford: Oxford University Press, 1958.
- Williams, Bernard. *Descartes: The Project of Pure Enquiry*. London: Penguin Books, 1978.
- . *Essays and Reviews, 1959–2002*. Princeton: Princeton University Press, 2014.
- . "Fictions, Philosophy, and Truth." *Profession*, January 1, 2003, 37–43.

———. *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Edited by Geoffrey Hawthorn. Princeton: Princeton University Press, 2008.

———. *Making Sense of Humanity and Other Philosophical Papers 1982–1993*. Cambridge: Cambridge University Press, 1995.

———. *Morality: An Introduction to Ethics*. London: Penguin, 1972.

———. *Morality: An Introduction to Ethics*. Reissue. Canto, 1993.

———. *Moral Luck*. Cambridge: Cambridge University Press, 1981.

———. *Philosophy as a Humanistic Discipline*. Princeton: Princeton University Press, 2008.

———. *Problems of the Self*. Cambridge: Cambridge University Press, 1976.

———. “Ryle Remembered.” *London Review of Books*, November 22, 1979.

———. “Seminar with Bernard Williams.” *Ethical Perspectives* 6, no. 3 (1999): 243–65.

———. *Shame and Necessity*. Berkeley: University of California Press, 1993.

———. *The Sense of the Past: Essays in the History of Philosophy*. Princeton: Princeton University Press, 2006.

———. *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press, 2002.

Williams, Bernard, and A. W. Moore. *Ethics and the Limits of Philosophy (with a Commentary on the Text by A. W. Moore)*. Abingdon: Routledge, 2006.

Williams, Bernard, and Michael Tanner. "Comment on 'Contraception and Chastity.'" *The Human World*, 1972.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Basil Blackwell, 1953.

Wolf, Raphael. "Socratic Authority." *Archiv Für Geschichte Der Philosophie* 90, no. 1 (2008): 1–38.

Wright, Crispin. "Self-Knowledge: The Wittgensteinian Legacy." In *Knowing Our Own Minds*, edited by Crispin Wright, Barry C. Smith, and Cynthia Macdonald. Oxford: Clarendon Press, 1998.