

# Two Levels of Moral Thinking<sup>1</sup>

Daniel Star

## I. Introduction

The purpose of this paper is to introduce a two level account of moral thinking that, unlike other accounts, does justice to three very plausible propositions that seem to form an inconsistent triad:

- (1) People can be morally virtuous without the aid of philosophy.
- (2) Morally virtuous people non-accidentally act for good reasons, and work out what it is that they ought to do on the basis of considering such reasons.
- (3) Philosophers engaged in the project of normative ethics are *not* wasting their time when they search after highly general moral principles which could not be discovered or be justifiably accepted through non-philosophical thinking, and which specify the good reasons that virtuous people act on, as well as provide a criterion or criteria for determining what it is that people ought to do.

The first of these claims is attractive because it is important to avoid an odious philosophical elitism which would have it that only philosophers can hope to be good people.<sup>2</sup> The second is a highly attractive idea that goes back to Aristotle: virtuous people are

---

<sup>1</sup> I would like to thank Mark Timmons and the originally anonymous referees, Michael Ridge and Christian Miller, for their very useful feedback, as well as all of the people in the audience at the *First Annual Arizona Workshop in Normative Ethics*, and in audiences elsewhere, who provided me with suggestions and criticisms.

<sup>2</sup> Despite much recent skepticism regarding the virtues, which has its origins especially in certain findings in social psychology, the ideal of virtue remains attractive. I believe that, at most, this skepticism undermines certain traditional accounts of the virtues, rather than contemporary theories that do not set the bar for moral virtue too high, and that appreciate that the virtues are commonly fragmented, fragile, and lacking in the type

not unaware of the moral reasons for which they act, and are not only luckily getting things right; rather, a virtuous person does the right thing for the right reason. The third claim seems highly attractive because normative ethics appears to be a valuable research program where we are making genuine, if slow, progress in articulating and defending fundamental moral principles which are not only such that we could not have begun by thinking of them as obviously true, but which clearly possess a degree of intellectual sophistication or complexity that is only possible to achieve and place trust in after extended philosophical reflection.<sup>3</sup>

These three highly plausible claims appear to give rise to an inconsistency because it seems that if (1) and (2) are true then (3) must be false: either the virtuous cannot non-accidentally act on genuine moral reasons – because we will only know these reasons when we do philosophy and discover the non-obvious principles that tell us what our genuine moral reasons for action are – or we do not need philosophy to discover the correct moral principles, for the virtuous can already know them.<sup>4</sup>

---

of unity that the traditional doctrine of the unity of the virtues supposes them to possess (see Adams 2006, for an excellent example of such a contemporary theory).

<sup>3</sup> I do not mean to deny that it is possible to stumble across sophisticated moral principles and understand them; however, what we look for in normative ethics is not mere acquaintance with some such principles, but good reasons to believe them. And I am assuming throughout that it is not necessary for us to be concerned with cases of people coming to know fundamental moral principles through testimony, since such cases, assuming they are possible (some philosophers have argued that it is a special feature of *moral* knowledge that this is not possible), would all require that *some* people know the fundamental moral principles in a way that is not based on testimony.

<sup>4</sup> Another option is to deny that there are any true and non-obvious moral principles waiting to be discovered, which is to deny (3) (see, for example, Dancy 2006). I will say something about particularism in the last section of the paper. Not everyone will think that the apparent inconsistency of the three central claims constitutes a problem, because not everyone will find all three claims attractive, but I hope a great many people will do so. It is a very interesting question how W. D. Ross's pluralist deontology (as outlined in Ross 2003, 16-64), which

In order to make good on and reconcile all three of these claims, I will suggest that we need to adopt a particular way of thinking about virtue, as well as a particular two-level account of reasons. Each of these is independently attractive, quite apart from the role they can together play in avoiding the problem that I have just described. Although this paper is intended to be schematic – leaving much detailed work still to be done – it should at least become clear here how adopting these accounts of virtue and reasons might enable us to avoid the apparent inconsistency in the above triad of claims, and also put us in a good position to identify which types of moral statements are good candidates for knowledge prior to philosophical reflection, and which are good candidates for knowledge only after philosophical reflection. I will end by tentatively exploring the possibility that the account of reasons I favor might be well placed to capture the grain of truth in particularism, while the account of virtue I favor might be well placed to capture the grain of truth in moral relativism.

Throughout, I intend to be completely neutral with respect to debates between the main theoretical options in normative ethics. My guiding belief is simply that, given the state of contemporary ethics, we have good reasons to think that whether the correct moral theory is consequentialist or not (deontological or not, etc.), it will be a theory that is intellectually

---

purposely stays very close to common sense morality, should be interpreted in relation to (3), but I do not have space to provide a completely adequate response to that question here. Suffice to say: one might very naturally take Ross to be denying (3), inasmuch as it can seem like he is not offering up new (justification for) moral principles at all, so long as one focuses on: (i) his claim that it is impossible to provide principles that specify what we ought to do when *prima facie* duties conflict and (ii) his descriptions of particular *prima facie* duties, which are such that they correspond very closely moral duties or reasons recognized by ordinary, non-philosophical moral thinking. On such a reading, there is not much left for us to do in normative ethics, in relation to questions about the fundamental principles of morality, at least, and (1) and (2) seem straightforwardly correct. This is not to say that deontological theories that take their inspiration from Ross (or someone inspired by much of the rich abstract theorizing that one also finds in Ross's work) need accept (i) and (ii), nor need necessarily reject (3).

sophisticated and non-obvious, from a non-philosophical perspective, in many of its fundamental details.<sup>5</sup>

## 2. Hare and Williams on Two Levels of Moral Thinking

A good place to begin is with an interesting exchange between R. M. Hare and Bernard Williams concerning the account of two levels of moral thinking outlined in Hare's *Moral Thinking* (1981). Hare begins this book by distinguishing between an 'intuitive' level and a 'critical' level of moral thinking. On the first level, we are said to follow various simple *prima facie* rules, confining our reasoning to how these best fit the everyday situations that we encounter. On the second level, it is said to be appropriate to take an impartial utilitarian perspective and assess the *prima facie* rules themselves, and revise them when necessary, as well as attempt to work out what to do when they conflict or otherwise provide inadequate guidance. In order to make good decisions about which rules to endorse on the critical level (rules that may sometimes be very limited in scope), Hare claims it is necessary to survey the preferences of every person that will be affected by particular actions, placing oneself in the position of each of these people in turn.

Hare recognizes that this is an ideal that we can only ever hope to imperfectly approach, rather than completely realize. He distinguishes between the ideal types of 'proles' and 'archangels'; the former make moral decisions only on an intuitive level – if they makes good decisions in a reliable fashion, this is only because they have received a good education from others who are more critical – while the latter only ever make decisions that are based

---

<sup>5</sup> One reason I think this is that I believe there is growing consensus that the correct theory will not be *too* distant from common sense morality in its prescriptions. I believe hedonistic act-utilitarianism is too distant from common sense morality to end up being a theory that any of us could end up believing with sufficient epistemic justification, since there is no good epistemic route from here to there.

on an accurate view of all of the consequences of any of the actions that are available to them. Actual human beings all fall somewhere in between these two extremes (Hare 1981, 45).

Williams expressed a number of concerns about this account of moral thinking in “The Structure of Hare’s Theory” (1988). Noting that Hare’s account of the two levels of moral thinking is somewhat like Sidgwick’s, Williams contends that it improves upon Sidgwick’s account by avoiding the arguably odious ‘Government House utilitarianism’ that would have it that there are two classes of people – an elite class who apply the correct ethical theory to particular practical issues, and a class of commoners who are (best kept) ignorant of the correct theory (Williams 1988, 188). Nonetheless, Hare is said to inherit from Sidgwick a fatal tension in his account of the two levels of moral thinking: the intuitive level is meant to correspond to ordinary non-philosophical moral judgments; however, “*the more the theory represents the intuitive reactions as merely superficial, provisional, and instrumental, the fewer appearances it saves*: it does not explain what people do feel and think, but suggests something else in the same area that they might usefully feel and think” (1988, 190, my italics). This sounds right, but it is not obvious what the best approach to saving the appearances is once we give up on Hare’s approach; I attempt to provide an alternative approach below. In any case, Williams argues that ordinary agents are not generally able to adopt the position of an archangel, even to the limited extent Hare contends that they are, and that ordinary agents are not generally inclined to accept the sometimes radical revisions to the *prima facie* rules that they hold dear which utilitarianism would have them undergo.

I believe Hare’s failure to give an account of the intuitive level of moral thinking that actually lines up with and appropriately explains our ordinary moral judgments is connected to two other closely related aspects of Hare’s project: his explicit rejection of any reliance on particular substantive moral judgments in his basic methodology, in favor of instead appealing only to formal conditions said to be required for the proper use of moral language; and his failure to provide any space within his project for legitimate claims about moral

knowledge, despite he himself suggesting his project has knowledge as its aim. First, with respect to his methodology, Hare (1981, 12) writes:

The appeal to moral intuitions will never do as a basis for a moral system. It is certainly possible, as some thinkers... have done, to collect all the moral opinions of which they and their contemporaries feel most sure, find some relatively simple method or apparatus which can be represented... as generating all these opinions; and then pronounce that that is the moral system we must acknowledge to be the correct one. But they have absolutely no authority for this claim beyond the original convictions, for which no ground or argument was given. The 'equilibrium' they have reached is one which might have been generated by prejudice, and no amount of reflection can make that a solid basis for morality.

For Hare, utilitarianism itself is to be justified *not* through reflection on our moral intuitions, but through observing, and thinking about what follows from features that are *necessary* for the proper use of moral language (i.e. *universalizability* and *prescriptivity*). It seems to most of us now that such formal linguistic conditions can not play the substantive normative role that Hare wanted them to play (I take it that this is the general consensus). Hare's attempt to ground the fundamental principle of normative ethics on prescriptivity and universalizability alone, while heroic, can be challenged both with respect to the relevant claims about what the necessary features of moral language are, and, more crucially than this, with respect to the steps that lead from these linguistic claims to the adoption of the utilitarian principle, in particular.

In any case, Hare's rejection of any reliance on the substantive moral judgments that make up our 'original convictions' takes for granted that these judgments are likely to have been generated by prejudice (to be fair, he says "might have been generated by prejudice", but the tone of the above quotation is one of damning skepticism). This is already to cast first-level moral thinking in a very weak role indeed, and the argument against (reflective) 'equilibrium' (which clearly has Rawls in mind) ignores the possibility that the process that

leads to such equilibrium might be better thought of as one that builds on knowledge, rather than on mere opinions.

It is telling that the epigraph to Hare's book is the following quotation from Plato (*Meno* 98b; Hare 1981, 1): "And truly I too speak as one who knows not – only guesses. But that there is a difference in kind between right opinion and knowledge, this, it seems to me, I do *not* guess; but of the few things, if any, that I would claim to know, this is one." This is an intriguing choice of epigraph because Hare, in fact, has very little to say about knowledge in his book; in particular, there does not seem to be any room in the theory outlined therein for *moral* knowledge. When he does speak of knowledge he is concerned with knowledge of people's preferences (and he then speaks of "...the sense of 'know' that moral thinking demands", 1981, 96), not with knowledge of what people ought to do. We are expected to leave behind our ordinary moral opinions and come to see that prescriptivism is correct, and that the most sensible prescription (one that fully appreciates the universalizability feature of moral language) is ultimately going to be a purely utilitarian *prescription*, which is not truth-apt.<sup>6</sup> That's *all* utilitarianism can be, since Hare's rejection of descriptivism leaves us no room

---

<sup>6</sup> Michael Ridge has pointed out to me that in saying this, I might be accused of misstating Hare's view, since in "Some Confusions about Subjectivity", originally published in 1976, Hare writes: "We can use the words 'right' and 'wrong' when speaking of moral statements that people have made. Some opponents of non-descriptivism have made a great deal of play with this fact, as they have with the similar fact that we also use the words 'true' and 'false', in certain contexts, of moral statements. But there is no reason why a non-descriptivist should not readily admit that these words are used in speaking of moral statements, provided at any rate that he is prepared, as most non-descriptivists have been since Stevenson, to allow that moral statements do have, as *one* element in their meaning, what is usually called 'descriptive meaning' ... For it may be this element to which we are adverting when we call a moral statement true or false." (Hare 1989, 26). Ridge believes Hare and Stevenson here anticipated certain elements of quasi-realism. That may be so, and quasi-realists may certainly have sophisticated things to say about moral knowledge (I see no *obvious* reason to think my own account of moral thinking would be inconsistent with such sophisticated quasi-realism). However, to say, as Hare does, that just part of a moral statement is truth-apt is a long way from saying that the whole of a moral statement is truth-apt, and knowing the descriptive meaning component of a moral statement (on the assumption that one is

to claim that we could ever come to know any moral claims, let alone know that we ought to (adopt *prima facie* rules that) impartially maximize utility.

Hare's Platonic rejection of any reliance on ordinary, substantive moral judgments (mere "opinion") goes hand in hand with his adoption of a normative ethical theory that conflicts with such judgments to too great an extent. I would like to suggest at this point that it is only by construing a class of first-level judgments as examples of unreflective knowledge that we can both make sense of the appearances, so far as ordinary agents are concerned (Star 2008), and make sense of the project of normative ethics, since this project can be most plausibly construed as one that is guided by an ideal of knowledge of fundamental moral principles that builds on pre-philosophical knowledge (if we assume that a purely coherentist account of epistemic justification is not an option, then it is important that we *start* with some moral knowledge). Hare's particular failure should not lead us to conclude that normative ethics – construed as a philosophical enterprise that aims to discover highly general and non-obvious explanatory moral principles – is a hopeless project (*contra* Williams).

Various moral theories, whether consequentialist or deontological, may be compatible with those of our first-level moral judgments that are good candidates for knowledge, insofar as they are compatible with certain core judgments regarding right and wrong action (e.g. that it is always wrong to torture someone merely for fun), and certain core judgments regarding normative reasons (e.g. that there is always some reason to be kind to people wherever possible, and that there is normally some reason not to lie). It may be that

---

attempting to follow Hare's outmoded way of distinguishing between descriptive and prescriptive meanings) is a long way from knowing the moral statement as a whole – the last may be impossible if part of the statement is essentially prescriptive and a different part of the statement is essentially descriptive. In any case, it is not at all clear what the descriptive meaning component of the fundamental utilitarian principle actually is, or whether that principle would even have very much descriptive content, assuming prescriptivism were correct.



utilitarianism does not fall into this category – that is, it may be the case that Williams is right that utilitarianism is too revisionary – but this should not deter us from developing better consequentialist or deontological theories that might better fit together with our most fundamental first-level moral judgments (and, ultimately, coming to rationally judge that one particular theory is correct).

### 3. Two Levels of Normative Reasons

I believe that in order to successfully develop a more satisfactory two-level account of moral thinking, it is necessary to: (1) provide an account of reasons that divides reasons into two types: those that the folk can non-accidentally follow, and those that philosophers can discover (and, ideally, come to know); (2) provide an adequate explanation of how these two types of reasons are related; and (3) provide a plausible account of virtue according to which being minimally virtuous will depend only on first-level thinking, i.e. thinking about and responding to first-level reasons. Achieving the first task will enable us to avoid odious elitism, and establish a more attractive division of labor between the folk and moral philosophers, and achieving the third task will place virtue squarely with the folk. The second task is also crucial, because it is one thing to posit two categories of reasons, and quite another to avoid the appearance of simply trying to have one's cake and eat it too. In this section, I will attempt to undertake the first two tasks, and I will attempt to undertake the third task in the next section.

The new account of reasons that I hope will enable us to achieve these tasks has at its heart a unified and informative analysis of normative reasons that has been argued for by myself and Stephen Kearns elsewhere (Kearns and Star 2008, 2009, 2010). We call this analysis *reasons as evidence*. It can be stated as follows.

R: Necessarily, a fact F is a reason for an agent A to  $\phi$  *if and only if* F is evidence that A ought to  $\phi$ .

This analysis of reasons explains what it is for a fact to be a reason in terms of that fact being evidence for the truth of an ought proposition. In the case of reasons for belief, “ $\phi$ ” is to be replaced by “believe P”, and with respect to normal, purely non-pragmatic reasons for belief the fact that is evidence that I ought to believe P is evidence that I ought to believe P just *because* it is evidence that P.

An example might aid in understanding R, as it applies to reasons for action. Imagine that you are hurrying to meet a friend, Jed, who you promised to meet in a few minutes for coffee, and you come across a sick stranger who needs your assistance to get to a hospital, and it is apparent to you that providing such assistance will take quite a bit of time (so it is not possible to both help the stranger and keep your promise to Jed). A natural way of describing your decision to help the stranger (if that is the decision you would make) is to say that you recognize that there is a reason to meet your friend and a reason to help the stranger, and you judge that the reason to help the stranger is a stronger reason than the reason to meet your friend. Another natural way of describing your situation is to say that you recognize that there is evidence that you ought to meet your friend and evidence that you ought to help the stranger, and you judge that the evidence that you ought to help the stranger is stronger than the evidence that you ought to meet your friend. *Reasons as evidence* has it that the relevant claims about evidence and reasons are (extensionally) equivalent.

We can speak, as I just did, of it being the case that there “is” a reason, just as we can speak of it being the case that there “is” evidence. Similarly, we can speak of it being the case that a person “has” a reason, just as we can speak of it being the case that a person “has” evidence. If a stranger outside my front door urgently needs my help but I am inside and unaware of that fact, there is still a reason for me to help the stranger (there is evidence that I

ought to help the stranger), but this is not a reason (evidence) that I possess. Of course, there may be constraints on which facts can count as unpossessed evidence, and these constraints will also be constraints on unpossessed reasons. Plausibly, a fact  $F$  can only be evidence for me that  $P$  if an *idealized* counterpart of myself would have this evidence (i.e. be aware of  $F$  and recognize that it is evidence that  $P$ ). Instead of  $R$ , one could instead choose to initially focus on the following claim:  $A$  has a reason to  $\phi$  and that reason is  $F$  *iff*  $A$  has evidence that he ought to  $\phi$  and that evidence is  $F$ . One could then define what it is to *be* evidence that one ought to  $\phi$  as follows:  $F$  is evidence that  $A$  ought to  $\phi$  *iff* an appropriately idealized counterpart of  $A$  would have evidence that he ought to  $\phi$ , and that evidence would include  $F$ . Finally, one could specify what it is for there to *be* a reason (i.e. derive  $R$  from the preceding claims):  $F$  is a reason for  $A$  to  $\phi$  *iff*  $F$  is evidence that  $A$  ought to  $\phi$ . I will not attempt to say anything here about what kind of idealization is most appropriate for this account of reasons, but that is certainly something that would be worth thinking about.

According to  $R$ , all reasons for action are evidence that I ought to do something in particular. Going beyond this core analysis, we might additionally suppose that some of the facts that are reasons *also* play an ultimate right-making role (or wrong-making role); they *make it the case that* particular acts are right (or wrong). These *ultimate* reasons would be the reasons that we aim to specify in normative ethical theories, along with universal principles concerning what we *ought* to do.

Suppose, for the sake of a simple example, that the correct moral theory is hedonistic utilitarianism (needless to say, I don't actually think this is the correct moral theory). This theory would have it that the only *ultimate* reasons are facts about pain and pleasure. If an act would increase pleasure in the world then the fact that this act would increase pleasure is an *ultimate* reason to do it, while if an act would increase pain in the world then the fact that this act would increase pain is an *ultimate* reason not to do it. Furthermore, one ought to do those acts that increase the balance of pleasure over pain (according to the theory). Now consider the fact that a particular act is a lie. This normative theory says nothing about facts

that are lies. However, the fact that this act is a lie is evidence that one ought not to do it, and is thus also a *derivative* reason not to do it. What ultimately makes it (reliable) evidence that one ought not to lie, assuming it is, is that it is generally true that lying leads to an increase in pain.

Of course, discovering *ultimate* reasons will be no simple task, but we are all familiar with *derivative* reasons; that is, on the present account of reasons, mundane evidence concerning what we ought to do (and ought not do). Plausibly, it is by thinking carefully and systematically about this mundane evidence (amongst other things) that philosophers are able to develop good ethical theories that provide us with deep explanations of what is right and wrong. The direction of explanation seems to be the reverse of the direction of discovery – ultimate reasons explain derivative reasons, but it is only through first encountering derivative reasons that we are able to discover ultimate reasons. Furthermore, we may be in a position to know what our (good) evidence is; if that is so, we are in a position to *knowingly* responding to our (derivative) reasons. This, I am now well-placed to suggest, is the position that the (minimally) virtuous find themselves in.

To illustrate using the above example: on the assumption that utilitarianism is true (contrary to what I actually think), a virtuous agent might justifiably recognize that there are reasons not to lie, even when they have never heard of, let alone have thought carefully about, utilitarianism; they would be able to do this because they are able to recognize ordinary evidence that they ought not lie (the fact that it upsets people when they discover that they have been told a lie, or the fact that they themselves feel bad when they have been lied to, to use a couple of superficial examples). Such evidence might itself be an input into a subsequent attempt to discover the correct normative ethical theory, even though it would not ultimately play a role in the spelling out of that theory.

I believe that only *reasons as evidence* can explain how it is that the direction of explanation can be the reverse of the direction of discovery in the way I have just described. At least, this seems to be the only account of reasons that might hope to do so in a way that is

true to the appearances, so far as the moral reasons we encounter in our ordinary lives are concerned, at the same time as providing room for there to be the ultimate reasons that sophisticated first-order moral theories would posit. In particular, it can hope to do this in a way that is promising when we think about moral epistemology: a crucial challenge for alternative accounts of reasons is to explain how it is that we can know, or be justified in following, derivative reasons, without possessing any knowledge of ultimate reasons. *Reasons as evidence* does not hide the key to this puzzle: we can start off being aware of reasons, in virtue of the fact that we start off with evidence concerning what it is that we ought to do, and these reasons, as evidence, come ready to also lead us to deeper knowledge through reflection.

Admittedly, I have not surveyed alternative accounts of reasons here. Here I can only quickly provide a criticism that will apply to a large and popular class of such accounts. A popular idea is to think of a normative reason as being at least part of an *explanation*: it might, for instance, be part of an explanation of why one ought to do a particular act, or an explanation of why an act is desirable.<sup>7</sup> The problem with accounts of this form is that parts of explanations are not well-suited to play the role in practical reasoning that normative reasons play. I take it that two truisms about normative reasons are: (1) that they are typically weighed against each other; and (2) that they are typically inputs, rather than outputs in pieces of practical reasoning. In relation to both (1) and (2), reasons contrast with oughts –

---

<sup>7</sup> John Broome defends a sophisticated version of the first view (for extensive commentary on Broome, see Kearns and Star 2008), and Sean McKeever and Michael Ridge are developing a version of the second view. It is also quite common to think of reasons as right-makers (or wrong-makers). “Right-maker” is a term of art: it can either refer to one side of some kind of explanatory relation, in which case the criticism above applies (unless one also accepts reasons as evidence), or to one side of some kind of metaphysical relation, e.g. a constitution relation. I can not do justice to the second option here, but my concern with it would be partly metaphysical and partly epistemological: how are we to understand the constitution relation in question, and how do right-makers, on this understanding, admit of both derivative and ultimate kinds (keeping fixed the desideratum that right-makers of the derivative kind come first in the order of discovery)?

both of these claims are not true of oughts, since judgments about what one ought to do typically come as conclusions in practical reasoning. In any case, reasons are typically upstream of oughts. Explanans, however, are typically downstream of explanandum, in the sense that one starts with something to be explained and then looks for ways in which it might be explained. In practical reasoning, one cannot *start* with explanations of something one ought to do, because by paying attention to reasons one is trying to work out what one ought to do! On the other hand, *reasons as evidence* fits together very well with (1) and (2).

One might be tempted to respond to this criticism of explanation-based accounts of normative reasons by saying the following: the facts that are reasons in virtue of playing an explanatory role can play the right kind of role in practical reasoning without being appreciated or understood *as* (parts of) explanations by agents engaged in practical reasoning.<sup>8</sup> However, I see little reason to accept that this is a live possibility, since the only way this alternative picture could be correct is if the folk were able to systematically identify the right facts *as* reasons *without* thinking of them as playing an explanatory role, where the right facts happen to be those that would also be independently identified *as* explanations by philosophers, and we have yet to be told by what means the folk are able to do this (until we are provided with a plausible story as to how this could occur, such a strong correlation seems miraculous).

Suppose that I am right that *reasons as evidence* is the best account of reasons to place at the heart of a two-level approach to understanding reasons. It might be objected that this is neither here nor there, because all this talk of there being two levels of reasons – one for the ordinary virtuous person and one for the philosopher – already hopelessly clashes with the non-philosophical commitments of ordinary moral agents; ordinary agents will wish to deny that there *is* a deeper explanation of our moral reasons waiting to be provided.

---

<sup>8</sup> Michael Ridge responded to the criticism in this way when I raised it at the *Seventh Annual Metaethics Workshop*.

Contrary to what some have claimed,<sup>9</sup> I do not believe commonsense morality contains any such commitment to there being no deeper philosophical explanations of moral truths.

I agree that commonsense morality would have us save many of the appearances, and thus might well be thought to conflict with *some* moral theories (act-utilitarianism, for example). However, this claim should not be confused with the claim that commonsense morality rules out the possibility that deeper explanations of moral truths might be provided. On the contrary, I think the folk are committed to thinking of the ordinary, virtuous person as being epistemically modest about many moral truths (including any possible fundamental principle of morality); and if I came to think I was wrong about this being something the folk are committed to, I would instead say that immodest commitments are not worth saving.

An analogy may be helpful here: a general ontological theory in metaphysics that would have it that *all* that exists are quantum particles or forces would clash with commonsense ontological commitments; however, there is no reason to think that such commitments would conflict with a general ontological theory that would have it that the most fundamental constituents of reality are quantum particles or forces, and that would also contend that it is a useful and important project to attempt to establish how it is that the objects we are ordinarily more aware of (chairs, people etc.) manage to exist in ways that are grounded in such fundamental facts.

#### 4. The Virtuous

I believe that a satisfactory account of virtue needs to meet three desiderata: (1) The acts that the virtuous typically perform are taken to be appropriate responses to facts that ultimately explain the rightness of the acts themselves (i.e. explain, for each right act they perform, why they *ought* to perform this act); (2) It is admitted that the virtuous can be, and

---

<sup>9</sup> See Nick Zangwill's paper in the present volume.

often are, *ignorant* of such ultimate explanations, and the facts such explanations depend on (at the very least, they can be ignorant that these facts are reasons); and (3) The virtuous always act *for* genuine normative reasons (i.e. they are responding to normative reasons when they act). The second and third desiderata correspond to two of the three claims that made up our initial apparently inconsistent triad. The first desideratum adds something quite new and distinctive; it asks us to think of the virtuous as responding, in some relevant sense, to ultimate right-makers (and wrong-makers), and not just think of the virtuous as responding to derivative reasons. It might seem that it is difficult to develop an account of virtue that meets all three desiderata, just as it might have initially seemed difficult to avoid inconsistency without denying any of the three highly plausible claims we started with.

Here is a summary of the account of virtue that I think best meets these desiderata, and is also attractive in its own right. Rather than follow neo-Aristotelian virtue ethicists like Rosalind Hursthouse (1999), who construe virtuous agents as morally omniscient (whenever the relevant non-moral knowledge is in place and there is a determinate answer to the question “what ought to be done?”), we should think of them as agents who track and respond reasonably well to *pro tanto* derivative reasons, guided by knowledge of what is *normally* right and wrong. They must also possess knowledge of what they always ought (not) to do, all-things-considered, in a large range of paradigmatic cases (e.g. one always ought to help people in need when this imposes no significant cost on anyone, and it is always wrong to torture people merely for fun). We should think of virtuous agents as being very skilled at weighing up (derivative) reasons, but also as often lacking knowledge of true normative principles that would allow them to work out what they ought to do in very difficult cases, where the reasons they encounter are of a similar weight. Very plausibly, good agents possess the virtue of *normative modesty* – that is to say, when we think carefully about ordinary virtue, we see that the virtuous do not take themselves to have all the answers to moral questions (to do so would be a vice).



It is very tempting to think that in difficult cases, the virtuous could benefit from the provision of an ethical theory that specifies universal criteria for all-things-considered rightness and wrongness (i.e. tells them what they ought and ought not do), but also that their inability to bring to mind or justifiably endorse such a theory does not detract from their virtue. To be virtuous, one need not be maximally virtuous, after all.

This account, which is admittedly rather sketchy, meets the specified desiderata. Notice, in particular, how desideratum (1) is met since, on this account of virtue, the virtuous respond to derivative reasons, and the analysis of reasons that was provided above connects both derivative and ultimate reasons directly to ought facts. One can respond to ultimate reasons *indirectly*, via derivative reasons, while acting in the same way that one would act if responding to ultimate reasons directly.<sup>10</sup> To return to the earlier example: I can respond to the fact that an act would cause someone pain *indirectly* by responding *directly* to the fact that the act would be a lie, without for a moment taking into account the fact that it would

---

<sup>10</sup> When Mackie presents the famous argument from relativity (now often called ‘the argument from disagreement’), he also presents a “well-known counter to this argument” (1977, 37). His argument depends on first recognizing that there are many moral disagreements between people of different groups, and then accepting that the best explanation of this disagreement is that differences in moral codes are caused by differences in ways of life, and not the other way around. The counter to the argument that he considers is that there may be some principles which are recognized in all societies, and that many of the variations we see in moral codes are due to differing applications of these principles to differing circumstances. His reply to this objection is that “people judge that some things are good or right, and others are bad or wrong, not because – or at any rate not only because – they exemplify some general principle for which widespread implicit acceptance could be claimed, but because something about those things arouses certain responses immediately in them, though they would arouse radically and irresolvably different responses in others. ... the argument from relativity [thus] remains in full force” (1977, 37-8). The present approach to normative reasons and virtue provides a way of responding to this reply: ordinary virtuous people may not be directly aware of the deep principles that specify the features that always make certain acts right and certain acts wrong, but they may be responding to these features anyway, *via* responding to evidence concerning what they ought to do.

cause pain.<sup>11</sup> Furthermore, my response (in this case, choosing not to lie) would be the same regardless of whether I was well-placed to focus on facts about lying, or facts about pain, as reasons. To use a more positive example: I might take my reason to cheer you up when you are worried to be the fact that you are worried (this is what I think to myself), but it could be that by directly responding to the fact that you are worried I am indirectly, and appropriately, responding to the fact that you are in pain (or responding to the fact that your rational autonomy is being undermined, say).

Here is an objection to the preceding account of virtue: suppose Joe happens to believe that facts about the relative positions of stars in the Milky Way are evidence concerning what he ought to do, not just in relation to particular decisions he needs to make as a scientist in an observatory (for there are non-bizarre ways that the positions of stars could count as reasons for action), but, quite bizarrely, in relation to all of his moral decisions; furthermore, just suppose that, even more bizarrely (through some huge cosmic coincidence), these facts really are reliable evidence concerning what Joe ought to do whenever he makes any kind of practical decision whatsoever; now, notice how it seems quite incredible to think that Joe could count as a morally virtuous agent because he reliably tracks and responds to such facts when making practical decisions.<sup>12</sup> Although this may initially seem like a good objection, I believe it falls apart as soon as one clearly distinguishes

---

<sup>11</sup> Remember that this is meant to be a toy example. If the example seems like one where it would be easy for a virtuous person to actually directly attend to the ultimate reason in play, bear in mind that there are good reasons to think that the correct normative ethical theory will actually posit quite sophisticated facts when it comes to ultimate reasons. Perhaps the fact that an act would respect rational autonomy would be a good candidate for an ultimate reason that would better make the point here (since it is plausible that ordinary virtuous agents do not think thoughts about rational autonomy, but rather focus on considerations to do with honesty, kindness, courage etc.).

<sup>12</sup> I am grateful to Howard Nye for raising this objection at the *First Annual Arizona Workshop in Normative Ethics*.

between two versions of the imagined scenario: Joe may have formed his beliefs about what is and is not evidence concerning what he ought to do in either an epistemically *irresponsible* or epistemically *responsible* fashion. If Joe formed his beliefs about the relevant evidence in an epistemically irresponsible fashion (which seems much easier to imagine than the alternative), then we shouldn't think he is virtuous at all. To be sure, I didn't previously mention this as a condition for possessing moral virtue, but it now seems we need to add this condition to our account of virtue, and I don't see any reasons why we shouldn't.<sup>13</sup> On the other hand, if Joe somehow managed to form his beliefs about the relevant evidence in an epistemically responsible fashion, then I think we would properly view him as morally virtuous.

At this point, I am in a position to provide a summary statement of the new two level account of moral thinking. On a first level, moral thinking involves the collection and weighing up of basic forms of evidence concerning what one ought to do, i.e. an appreciation of and an ability to weigh derivative reasons. It seems that most of us already possess such evidence (or are close to possessing it), and may also possess certain virtues by responding well to certain types of evidence (e.g. the evidence that we classify under 'kindness'). On a second level, moral thinking involves essentially philosophical thinking that moves from paying attention to such evidence in a way that guides our actions to thinking about it systematically in order to postulate and, ideally, come to know ultimate explanations for the rightness or wrongness of particular acts, on this basis. Such thinking might also take us beyond minimal virtue in enabling us to determine what we ought to do in hard cases, where we previously only saw (derivative) reasons of roughly equal weight.

---

<sup>13</sup> In fact, I am happily heading in the direction of viewing most of the practical virtues as partly reducible to epistemic virtues (*partly* reducible because they also have an affective dimension). As I also mention below, I think *enkrasia* is one of the practical virtues that can't be understood in this reductive fashion.

## 5. Moral Particularism and Cultural Variability in the Virtues

Now that I have provided a way of resolving the apparent inconsistency that I began with, while not giving up on any of the three main claims, I will finish by suggesting that the proposed two level framework may have two additional attractive features. The accounts of virtue and reasons that I sketched above seem better placed than other accounts on offer when it comes to explaining some of the key intuitions that feed particularists, as well as some of the key intuitions that lead people to be tempted by moral relativism out of a desire to understand and appreciate cross-cultural variation in the virtues.

The account of reasons that I introduced above may be able to provide an explanation of many of the intuitions that motivate moral particularists; an explanation which might, in fact, lead us away from particularism. I am thinking, in particular, of the intuitions that Jonathan Dancy (2006) relies on when he promotes the holism of reasons thesis that he believes makes it impossible to articulate interesting moral principles. If we separate reasons out into our two classes of derivative and ultimate reasons, we may be in a good position to deny that there is holism on the level of ultimate right-makers and wrong-makers. Derivative reasons may have an holistic structure – and to the extent that this is so, particularists might be onto something important when it comes to the deliberations of the virtuous – but the particularist may be confusing intuitions that involve derivative reasons with the intuitions that we can come to have about ultimate reasons. Ultimate reasons may have an atomistic structure on the account I sketched above, since they *always* count in favor, or *always* count against, particular acts.<sup>14</sup>

---

<sup>14</sup> I am ignoring a complication here. McKeever and Ridge (2006) have argued that it is possible to reject particularism (about moral principles) while still maintaining that the structure of reasons is holistic through and through. While this might be true, merely as a point about the logical relations between the bare particularist position and the bare holist position, I believe holism is, in fact, explanatorily redundant once one rejects particularism. I discuss reasons for thinking this is so (that do *not* depend on *reasons as evidence*) in Star (2007).

Once again, it might help if we use an example. Very plausibly, there is *normally* a reason not to lie, and on many individual occasions one may correctly judge that there is a reason not to lie. Particularists like to point out that it is also very plausible to suppose that one can correctly judge that there is *no* reason not to lie when one is in a situation where one is playing a game with friends that everyone knows requires lying in order to win (to use one of Dancy's favorite examples). Examples such as this one are taken to be evidence in favor of the holism of reasons thesis, since such examples seem to suggest that reasons are not generally invariant, i.e. that they can change their 'valency' from situation to situation in a way that always depends on other features of particular situations. Yet, on the present two-level account of reasons, the ultimate wrong-making feature in cases where it is wrong to lie (e.g. the disrespect of another's rational faculties, or pain, or whatever features would be cited in the correct ethical theory) can instead be thought of as being part of an atomistic structure of ultimate reasons. There is no reason not to lie when playing a game, precisely because no ultimate wrong-making feature is present in such a case.

This last point seems perfectly generalizable. However, one might worry that since I claim that the virtuous do not need to be aware of ultimate reasons, an adequate explanation of why an ordinary virtuous agent might appropriately judge that there is no reason at all not to lie in a situation that involves a game that requires lying to win should not appeal to the absence of an ultimate reason. I think this worry can be adequately responded to as follows. What ultimately explains why there is no reason not to lie in the game case is the absence of an ultimate reason not to lie, but what explains how the agent can judge that he has no reason not to lie may be something quite different. Assuming reasons are evidence (in the exact manner specified above), we can explain 'silencing' on the level of *derivative* reasons by utilizing the notion of an *epistemic defeater*. The ordinary agent ordinarily takes the fact that an act will be a lie to be evidence that he ought not do it, but in certain cases this evidence can be defeated by other evidence that he possesses, such as strong evidence (or knowledge) that everybody has agreed to have fun by lying within the boundaries of a game.

The new account of reasons can also provide an explanation for why the strength of a derivative reason and the strength of an underlying ultimate reason do not need to be added together in those cases where an agent is aware of both reasons. One might worry that any highly promiscuous account of reasons for action, such as *Reasons as Evidence*, will lead to a counting problem, and perhaps then to an unattractive solution to this problem that supposes that some reasons have no weight at all (see McKeever and Ridge 2006, 132-3; and Star 2007). Fortunately, *Reasons as Evidence* can provide an explanation of why we can not always aggregate the strengths of reasons: some items of evidence are *independent*, while some items of evidence are *not independent*. It is not rational to simply add the strength of two dependent reasons together (assuming one has evidence that they are not independent). In becoming aware of ultimate reasons (through philosophical reflection, recall), agents become aware that derivative reasons are not independent of certain ultimate reasons. Although it is no easy task for epistemology to explain how we make good judgments about when some evidence is dependent on some other evidence, it is clear that we all do this very often when it comes to ordinary reasons for belief, so whatever explains how we can do this in non-controversial cases involving reasons for belief will also enable us to explain how we can do this when it comes to reasons for action.

Finally, let me briefly mention a way in which the new account of virtue might be further developed and utilized to capture intuitions concerning cultural variability and disagreement when it comes to the virtues. Consider a recent attempt to provide disentangling, semi-reductive analyses of *thick* virtue terms in terms of descriptive elements and *thin* moral terms. Elstein and Hurka suggest that courage (for example) can be analyzed in the following way: "...'act x is courageous' can be analyzed as something like 'x is good, and x involves an agent's accepting harm for himself for the sake of goods greater than the evil of that harm, where this property makes any act that has it good,' and where... the second 'good' is an embedded evaluation." (2009, 527). They suggest that it might be the case that all *thick* virtue terms can be analyzed into a core descriptive component, and a *thin* evaluative

component (reference to *good* is essential in the above analysis of courage). An alternative type of semi-reductive analysis to the one that Elstein and Hurka propose might start with the following thought: perhaps individual virtue terms are best analyzed by reference to evidence concerning what one ought to do (i.e. reasons for action) of various types, where the type is picked out by the descriptive (i.e. non-normative) component of the relevant thick virtue term.

By focusing on evidence, this way of analyzing particular virtue terms might be well-placed to capture intuitions about justifiable cultural variability.<sup>15</sup> For example, courage may consist in acting in ways that involve risking harm to oneself where there is evidence that one ought to accept the relevant risks. One possible precise analysis of courage is the following: to say that an act *x* is courageous is to say that *x* involved the agent risking harm to himself for reasons that the agent possessed, and that he rationally judged to be stronger than reasons that the agent also possessed not to risk such harm. Given *reasons as evidence*, this is equivalent to claiming that to say that an act is courageous is to say that *x* involved the agent risking harm to himself on the basis of evidence that he ought to do so that the agent possessed, and that he rationally judged to be stronger than the evidence that he ought not to risk such harm. This type of semi-reductive analysis is one that focuses on the normative, rather than the evaluative sphere. It seems a little simpler than the Elstein and Hurka style of

---

<sup>15</sup> An alternative would be to simply accept the Elstein and Hurka proposal for analyzing virtues, alongside *reasons as evidence*. In an earlier draft, I claimed that Elstein and Hurka's proposal might not be as well-placed to explain cultural variability in the virtues as a version of that proposal that incorporates *reasons as evidence*. Hurka subsequently convinced me that there is no particular reason, for all I have said, to think that this is so; after all, he and Elstein themselves discuss the fact that their analysis leaves space for a huge amount of disagreement about which things are, in fact, good, e.g. integrity "involves an agent sticking to a significantly good goal despite distractions and temptations", but there will much disagreement as to which goals are really good to stick to (2009, 526). I also do not mean to say at this stage that the alternative suggestion for a way to analyze virtues that I now favor is *better* placed for explaining cultural variation in the virtues. It certainly seems very well placed.

analysis, but that is not to say very much; I must postpone any attempt to decide which is the better type of analysis of virtue terms until a later occasion – at this stage, I simply offer it as an interesting alternative.

The relevant concept for each individual virtue may have fairly minimal content; on the present proposal, an act of mine can fall under the extension of the concept of courage if it involves a response to evidence concerning what I ought to do when it comes to confronting dangers to myself. Since such items of evidence also constitute reasons for action, individual virtues involve particular types of reasons for action. The minimal descriptive component of any virtue term may be all that unifies that virtue across very different cultures – this may be part of what explains why we still believe that certain acts *are* acts of courage (or kindness, or loyalty etc.) in another society, and why we think of courage (for example) as a trait that exists in other societies despite significant cultural differences. Considered as a character trait, rather than as a way of describing actions, courage may be understood as a fixed disposition to act on a certain type of evidence concerning what I ought to do (and, in the case of intellectual courage, what I ought to believe).

What one is justified in taking the reasons of courage to be, and what one is justified in taking to be appropriate responses to such reasons may very well be partly dependent on culturally specific practices. This need not commit one to saying that cultural practices are the *source* of evidence concerning what one ought to do; all one needs to commit oneself to is the highly plausible thought that such practices block, transmit, downplay, or foreground evidence. Some particular evidence that one ought to act in a risky manner may come to be possessed in one cultural context, but not in another, and similar things might be said for all of the other moral virtues.<sup>16</sup>

---

<sup>16</sup> I do not mean for this type of analysis to be extended to the epistemic virtues. There may be important exceptions even when we focus on practical virtues alone. In particular, the virtue of strength of will (*enkrasia*) seems of central importance to the ethical life, and this virtue might well be construed as being about sticking



## References

- Robert Adams, *A Theory of Virtue*, Oxford: Oxford University Press, 2006.
- Jonathan Dancy, *Ethics Without Principles*, Oxford: Oxford University Press, 2006.
- Daniel Y. Elstein and Thomas Hurka, "From Thick to Thin: Two Moral Reduction Plans," *Canadian Journal of Philosophy* 39:4, 2009, 515-36.
- R. M. Hare, *Moral Thinking*, Oxford: Clarendon Press, 1981.
- R. M. Hare, *Essays in Ethical Theory*, Oxford: Oxford University Press, 1989.
- Rosalind Hursthouse, *On Virtue Ethics*, Oxford: Oxford University Press, 1999.
- Stephen Kearns and Daniel Star, "Reasons: Evidence or Explanations?" *Ethics* 118:4, 2008, 31-56.
- Stephen Kearns and Daniel Star, "Reasons as Evidence," *Oxford Studies in Metaethics* 4, 2009, 215-42.
- Stephen Kearns and Daniel Star, "Weighing Reasons," ms., 2010.
- J. L. Mackie, *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin Books, 1977.
- Sean McKeever and Michael Ridge, *Principled Ethics*, Oxford: Oxford University Press, 2006.
- W. D. Ross, *The Right and the Good*, Oxford: Oxford University Press, 2003.
- Daniel Star, "Sean McKeever and Michael Ridge, *Principled Ethics*," *Notre Dame Philosophical Reviews*, 2007.

---

to one's weighing of evidence concerning what one ought to do. Such a construal, while it is friendly to the above approach, does not exactly follow the model of semi-reductive analysis provided above.

This is a draft. Please cite version published in *Oxford Studies in Normative Ethics* 1, 2011, 75-96 (email [dstar@bu.edu](mailto:dstar@bu.edu) for a copy).

Daniel Star, "Moral Knowledge, Epistemic Externalism, and Intuitionism," *Ratio* 21:3, 2008, 329-43.

Bernard Williams, "The Structure of Hare's Theory" in Douglas Seanor & N. Fotion (eds.), *Hare and Critics*, Oxford: Clarendon Press, 1988, 185-96.